#27

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE
BEFORE THE BOARD OF PATENT APPEALS AND INTERFERENCES

In re application of:			
Dane K. FISHER *et al.*		Art Unit:	1637
Appln. No.:	09/394,745	Examiner:	Young J. KIM
Filed:	September 15, 1999	Atty. Docket:	16517.001/38-21(15454)B
For:	***Nucleic Acid Molecules and Other Molecules Associated with Plants***		

APPELLANT'S BRIEF

Mail Stop Appeal Brief - Patents
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

This is an Appeal from the final rejection of pending claims 8-10 in the above-described patent application. A Notice of Appeal is filed concurrently herewith. Authorization to charge the official fees for this filing is given in the accompanying transmittal letter. *This brief is submitted in triplicate.*

In the event that extensions of time beyond those petitioned for herewith are necessary to prevent abandonment of this patent application, then such extensions of time are hereby petitioned. Applicants do not believe any additional fees are due in conjunction with this filing. However, if any fees under 37 C.F.R. §§ 1.16 or 1.17 are required in the present application, including any fees for extensions of time, authorization to charge such fees is given in the accompanying transmittal letter.

1. Real Party in Interest

The real party in interest is Monsanto Company, a Delaware corporation with offices at 800 North Lindbergh Boulevard, St. Louis, Missouri 63167.

2. Related Appeals and Interferences

Appellant previously filed an Appellant's Brief directed to the outstanding rejections of the pending claims under 35 U.S.C. §§ 101 and 112, first paragraph, on March 13, 2003. Appellant requests that the present appeal be consolidated with the appeal already pending in the present application.

3. Status of Claims

Claims 8-11 are pending. Claims 8-10 stand finally rejected because the Office has refused to examine the claimed invention. Applicants herein appeal all of the rejections of claims 8-10 that were not previously appealed, mainly those arising from refusing to examine claims 8-10 under the Commissioner's asserted authority to request restriction.

4. Status of Amendments

Applicants have not filed any amendments subsequent to the Final Office Action mailed September 11, 2002 (Paper No. 14) ("Final Action"), in this case. On January 10, 2003, Applicants filed a petition under 37 C.F.R. § 1.144 ("Petition") requesting the withdrawal of the restriction requirement to select a single combination of nucleotide sequences for examination. In response to the Petition, on February 14, 2003, the Office mailed a decision ("First Decision") denying Applicants' Petition. On April 14, 2003, Applicants filed a Request for Reconsideration of Applicants' Petition under 37 C.F.R. § 1.144 ("Request"), which was subsequently denied by the Office on May 12, 2003, in a second petition decision ("Second Decision"). It is the Applicants' position that the demand to select a single combination of nucleotides for examination and the subsequent acts of the Office, however denominated, constitutes a rejection; and that rejection has now been twice made, thereby being ripe for appeal.

5. Summary of Invention

The invention is directed to a microarray comprising a substrate with a surface comprising 10^3 nucleic acid molecules or more where at least 10% of the nucleic acid molecules are comprised of different sequences and at least about 250 nucleotide residues and complementary to a molecule comprising a sequence selected from a group. *See* specification at pages 59, line 25 through page 62, line 7.

6. Issues

The issue in this Appeal is whether the refusal to examine the invention embodied in claims 8-10 because of alleged undue burden constitutes an improper rejection, and particularly, whether the purported restriction requirement within a single claim (not between claims) and the insistence that the invention as claimed will not be examined constitutes an improper rejection.

7. Grouping of Claims

The patentability of claims 8-10 is addressed together in Sections 8.A through 8.D below. A copy of the currently pending claims, as presently pending in the application, is attached hereto as Appendix A. The complete text of claims 8-10, as originally filed in the Preliminary Amendment dated October 10, 2000 ("Preliminary Amendment"), are also attached hereto as Appendix B.

8. Argument

A. Summary of Appellant's Position

Applicants have disclosed and claimed a single invention. Applicants' invention is a microarray that can be modified by its user to select for specific nucleic acid molecules of interest in a given sample. Accordingly, the invention contemplates the ability to include or exclude any nucleic acid molecule comprising any one or more of the nucleic acid sequences that is a member of the group of nucleotide sequences set forth in claim 8. Admittedly, that puts

thousands of different combinations of nucleic acid molecules within the scope of claim 8, such is Applicants' invention – not a single combination. From the commencement of prosecution, the Office disregarded Applicants' claimed invention and required that Applicants select ***one*** combination of nucleic acid molecules for prosecution, ignoring the essence of the invention. In effect, the Office has refused to examine the invention as claimed. This refusal to examine what Applicants rightfully regard as their invention in accordance with 35 U.S.C. § 112, second paragraph, amounts to a rejection. *Ex parte Holt*, 214 U.S.P.Q. 381, 383 (B.P.A.I. 1982); *Appln. of Haas*, 486 F.2d 1053, 1056, 179 U.S.P.Q. 623, 625 (C.C.P.A. 1973).

That the Office purports to find authority in its restriction practice is irrelevant. It is well-settled law that a restriction requirement may only be issued ***between*** claims, not within a single claim. *Appln. of Weber*, 580 F.2d 455, 458-459, 198 U.S.P.Q. 328, 332 (C.C.P.A. 1978); *Appln. of Haas*, 580 F.2d 461, 464, 198 U.S.P.Q. 334, 336-337 (C.C.P.A. 1978); *Appln. of Haas*, 486 F.2d 1053, 1056. By requiring Applicants to elect a single combination for examination, the Office has improperly required restriction within a claim. No number of divisional or continuation applications directed to different combinations of the thousands of permissible combinations comprehended by the invention as originally claimed could ever amount to the whole of Applicants' invention. And even the full complement of individual applications directed to all those combinations would not effectively constitute the entirety of Applicants' invention, as the examination of each would not be the same as the examination of the integrated whole presented in claims 8-10.

The assertion by the Office that it would pose an undue search burden to perform a search on every combination envisioned by the claimed invention is not a justification for the improper refusal to examine the integrated whole. The Office is obliged to examine Applicants' invention and not some subset it would rather examine. Furthermore, the Office *should* examine Applicants' invention and eschew archaic procedures, policies and methods.

Despite outcries from the intellectual property community, the Office has resisted implementation of appropriate examination procedures in favor of overzealous and legally unsupportable restriction practice. First of all, the "undue burden" asserted by the Office has been imposed on the Office by itself because it clings to archaic search procedures. In any event, as the predecessor of the Federal Circuit has declared, "in drawing priorities between the Commissioner as administrator and the applicant as beneficiary of his statutory rights, we conclude that the statutory rights are paramount." *Appln. of Weber,* 580 F.2d 455, 458-459.

In short, the Office can examine claims 8-10 if it would do what the industry and academia (*see* Sec. 8.D below) do when it comes to searching. Its refusal to do so is the problem here, not Applicants' claims.

B. The Claimed Invention Was Never Examined by the Office

Claim 8 and its dependents were improperly subjected to restriction requirement, even under the guidelines set forth in M.P.E.P. § 803.04. Office Action mailed December 19, 2000 ("Restriction Requirement"), at page 3. This restriction requirement was based on a misconstruction of the nature of Applicants' claimed invention and, most importantly, it effectively denied Applicants their statutory right to have what they regard as the invention examined. 35 U.S.C. § 112, second paragraph. *See also Appln. of Haas:*

> we note that the board did attempt to 'go behind the bare words because of the possibility that a rejection in fact has been mislabeled as a withdrawal (or objection). . . That a de facto rejection had been made without statutory basis was also not considered. The withdrawal from any further consideration was instead classified as a 'refusal. . . to act on claims' in keeping with a restriction requirement and deemed a purely administrative matter. *Appln. of Haas,* 486 F.2d 1053, 1056.

Appln. of Weber:

> [a]n applicant is given, by the statute, the right to claim his invention with the limitations he regards as necessary to circumscribe that invention, with the proviso that the application comply with the requirements of s 112. *Appln. of Weber,* 580 F.2d 455, 458;

and *Appln. of Wolfrum*:

> [u]nder this provision of § 112, the scope of the subject matter is governed not by the examiner's conception of the 'invention' but by that 'which the applicant regards as his invention.' *Appln. of Wolfrum*, 486 F.2d 588, 591, 179 U.S.P.Q. 620, 622 (C.C.P.A. 1973).[1]

Applicants characterize the invention embodied in claim 8 as a microarray having a substrate with a surface comprising 10^3 nucleic acid molecules or more where at least 10% of the nucleic acid molecules (1) are different and (2) at least about 250 nucleotide residues and (3) complementary to a molecule having a sequence selected from a group. Petition at page 7. The invention includes thousands of collections and combinations of nucleic acid molecules for the skilled practitioner to select from when determining how best to screen a sample for the detection of a desired nucleic acid molecule or molecules. *See* Petition at pages 7-9. However, that invention was never examined by the Office.

Instead of recognizing that the invention claimed by Applicants is a microarray that has the ability to be modified by the user to select for a specific nucleic acid molecule or molecules of interest, the Office immediately divided the invention into subparts. Specifically, the Office required that Applicants elect a ***single*** combination and insisted that examination would be limited to this, and only this, combination. Restriction Requirement at page 3. Essentially, what the Office did was require that Applicants subject their invention to piecemeal examination, destroying the whole of the invention itself. Even if Applicants were able to bear the financial burden associated with filing and prosecuting thousands of applications,[2] each one directed to a

---

[1] This C.C.P.A. precedent is binding. *South Corp. v. U.S.*, 690 F.2d 1368, 1369, 215 U.S.P.Q. 657 ("[w]e hold that the holdings of our predecessor courts, the United States Court of Claims and the United States Court of Customs and Patent Appeals, announced by those courts before the close of business September 30, 1982, shall be binding as precedent in this court.")

[2] For example, at a minimum, if Applicants were to file 497 applications, each one directed to a microarray which must include a specific SEQ ID NO, as the Office has done here (*see* Sec, 8.D and footnote 5 *infra*), the cost to Applicants would be $ 372,750 in initial filing fees for the Office alone. Furthermore, even this expenditure would not allow Applicants' to have what they regard as the invention examined.

different combination that may be selected by the user of Applicants' claimed microarray, the end result would fall far short of accomplishing examination of the ***invention***. This is because the invention as claimed allows for the user to select which nucleic acid molecule or molecules to include or exclude from the substrate of the microarray. Piecemeal examination disregards the selectivity by the user which is embodied in Applicants' invention. *Appln. of Weber,* 580 F.2d 455, 458 ("[i]f, however, a single claim is required to be divided up and presented in several applications, that claim would never be considered on its merits. The totality of the resulting fragmentary claims would not necessarily be the equivalent of the original claim. Further, since the subgenera would be defined by the examiner rather than by the applicant, it is not inconceivable that a number of the fragments would not be described in the specification.")

The Office admits as much. Although in the First Decision the Office asserted that the restriction requirement applied to claim 8 and its dependents did not rewrite the claimed invention, in this same paper the Office states that

> Applicants may claim as many combinations as they wish. So long as the claimed combinations include SEQ ID NO: 5893, they are free of the prior art and no further search would be required to examine such claims.

First Decision at page 2. But all the combinations within Applicants' claims do not require the inclusion of SEQ ID NO: 5893, and the Office is not free to limit Applicants' invention to combinations which do include SEQ ID NO 5893. *Appln. of Weber,* 580 F.2d 455, 458 ("[i]f an applicant submits a number of claims, it may well be that pursuant to a proper restriction requirement, those claims will be dispersed to a number of applications. . . If, however, a single claim is required to be divided up and presented in several applications, that claim would never be considered on its merits.") Yet that is precisely what the Office has admitted it has done. The Second Decision further confirmed that the invention examined by the Office was not the same invention claimed by Applicants. The Second Decision states:

> First Applicants argue that the examiner, by forcing applicants to elect a single combination of sequences for examination, has narrowed the scope of the claims. ***This is correct, and that was the purpose of the restriction requirement.***

Second Decision at page 2 (emphasis added). Essentially, the Office has confirmed in these admissions that the invention claimed by Applicants was never examined, and was refused examination. Rather, the invention examined by the Office would require the user of Applicants' microarray to always include SEQ ID NO: 5893 on the microarray substrate, even if this sequence posed no interest to the researcher.

The convoluted prosecution history reinforces this point. From the very first office action, the Restriction Requirement, the Office and Applicants began talking past one another. The Restriction Requirement, relying on M.P.E.P. § 803.04, classifies each nucleic acid molecule as an "independent and distinct invention" and states that, as such, claim 8 was subject to a restriction requirement. Restriction Requirement at page 2. But Applicants are not claiming individual nucleic acid molecules that purportedly constitute independent and distinct inventions. They were and are claiming a microarray constituted by a combination of nucleic acid molecules that can be modified by its user to select for specific nucleic acid molecules of interest in a given sample. In any event, Applicants' response elected "the first 100 sequences in the Markush group."[3] Response to Restriction Requirement dated April 17, 2001 ("Response to Restriction Requirement"), at page 3.

The Office Action mailed March 18, 2002 ("First Office Action"), states that "the examination of SEQ ID Numbers will not go beyond the 100 SEQ ID Numbers." First Office Action at page 2. Yet, these first 100 SEQ ID Numbers were not examined properly,[4] rather, the

---

[3] Applicants stated:

> To the extent that this is not intended to be a further restriction requirement, but merely an election of species for examination, applicant elects the first 100 sequences in the Markush group.

Response to Restriction Requirement at page 3. This response was held by the Office as non-responsive and Applicants were required explicitly recite the 100 nucleic acid sequences selected.

[4] Accordingly, the examined invention no longer resembled Applicants' claimed invention, but rather a microarray that always included SEQ ID NO: 5893 – no matter what other sequences were included in the claim.

examination stopped as soon as one novel sequence was found in the combination, and it was presumed by the Office that Applicants' claims are limited to a combination which included that sequence found to be novel (*i.e.*, SEQ ID NO 5892). Even with the group of Applicants' claim 8 reciting only 100 SEQ ID Numbers, as it was later amended to do, thousands of combinations would still be included in the metes and bounds of the claimed invention. And those combinations were never examined – the examination improperly stopped once SEQ ID NO: 5893 was identified by the Examiner as novel over the prior art.

The Final Office Action mailed September 11, 2002 ("Final Action"), further accentuates the inappropriateness of the Office's position. The Final Action states that "Applicants could have elected all of the recited SEQ ID Numbers as the elected combination. However it was Applicants who have decided to elect the first 100 SEQ ID Numbers as the elected combination." Final Action at page 2 (emphasis added). In short, this statement affirms that fact that the claimed invention was never examined.

To the contrary, The only portion of Applicants' claimed invention that was actually examined by the Office were combinations in the microarray which include SEQ ID NO: 5893. In other words, the Office rewrote Applicants' claimed invention such that it always must include SEQ ID NO: 5893. But it is not what the Office regards as the invention that is tantamount. Rather, it is what the applicant sets forth under 35 U.S.C. § 112, second paragraph, as what he regards as the subject matter sought to be patented. *Appln. of Wolfrum*, 486 F.2d 588, 591 ("[u]nder this provision of § 112, the scope of the subject matter is governed not by the examiner's conception of the 'invention' but by that 'which the applicant regards as his invention.' ")

Applicants here are not claiming SEQ ID NO: 5893, either permanently embedded on a microarray substrate or in isolation. Applicants' invention is a microarray in which no single variation is characteristic of the claimed microarray, rather it is the ability to modify, substitute and select different collections or combinations of recited nucleic acid molecules, including the

ability to choose precisely which molecules to include or exclude from those listed in the recited group, that creates the essence of Applicants' invention. Applicants' invention is not a fixed microarray, but rather the invention provides the ability to vary the contents of a microarray within the parameters set forth in the claim. And as such, Applicants' claimed invention was never examined by the Office and this amounted to a rejection of claims 8-10.

It is apparent that the Office is intent on avoiding examination of Applicants' claimed invention and instead is seeking to find an easy way out by limiting the invention to an uncharacteristic trait (i.e., a single molecule) that guarantees novelty. However, it is the Office's obligation to examine for novelty what Applicants present as their invention, not to artificially inject novelty into a claim by treating the claim as if it were limited in a way the Office would like it to be limited.

C. The Restriction Requirement Amounts to a Rejection of the Claimed Invention

As previously mentioned, it is well-settled law that a restriction requirement may only be issued ***between*** claims, not within a claim directed to a ***single*** invention. *Appln. of Weber*, 580 F.2d 455, 458-459; *Appln. of Haas*, 580 F.2d 461, 464; *Appln. of Haas*, 486 F.2d 1053, 1056. The Office may not divide up a single invention for piecemeal examination.

Claim 8 and its dependents embrace a single invention. The restriction requirement divided that invention into single combinations which did not constitute the more generally claimed microarray. The Office relies on M.P.E.P. § 803.04 to classify each nucleic acid molecule within Applicants' invention as an "independent and distinct" invention, however, this classification fails to take into account the whole of the invention claimed. Applicants are not claiming nucleic acid molecules in isolation, nor are Applicants claiming a combination of molecules embedded on a substrate of a microarray which are characterized by the presence of one particular nucleic acid molecule (*i.e.,* SEQ ID NO 5893). Rather, the disclosed and claimed invention is a microarray that allows one to efficiently analyze large amounts of nucleotide

sequences for a target sequence or a fragment of that sequence, and to vary the parameters of that search based on available nucleic acid molecules comprising a selected assortment of nucleic acid sequences.

The Office may not rely on an articulated policy to arbitrarily apply restriction practice where it is not proper. *Ex parte Holt*, 214 U.S.P.Q. 381, 384. By imposing restriction within the invention encompassed by claim 8 and its dependents, the Office removed from consideration the selectivity crux of Applicants' invention. In addition, because the restriction requirement divided the invention claimed, no matter how many applications are filed directed to particular combinations of the claimed microarray, Applicants are precluded from presenting their integrated invention in another application. *Cf. Appln. of Hengehold,* n.11, 440 F.2d 1395, 1404, 169 U.S.P.Q. 473, 480, 58 C.C.P.A. 1099, 1110 (C.C.P.A. 1971). The restriction requirement did more than just divide Applicants' invention, it destroyed the invention as a whole.

Furthermore, the overzealous use of restriction practice, particularly as applied to biotechnology inventions, effectively denies Applicants any recourse in the courts. The Office here has demonstrated its willingness to apply § 803.04 indiscriminately, without regard to the invention actually claimed. In doing so, the Office may gain what it might perceive as a benefit in refusing to examine the invention claimed by Applicants, but the Office does so only by simultaneously purporting to prevent an appeal of the issue to a higher authority (*i.e.,* by calling its refusal to examine the claimed invention, which amounts to a rejection, a restriction). It is unequivocal that the Office refused to examine the invention claimed by Applicants and attempted to divide the microarray into pieces that could under no circumstances amount to the whole of the invention, including the selectivity feature. Thus, under the guise of restriction practice, the Office has in fact rejected Applicants' invention.

At the same time, because restriction practice is a petitionable matter, no recourse may be left for Applicants to immediately appeal the *de facto* rejection of the claimed invention. The Office is thus able to continue, unfettered, a policy of rejecting Applicants' invention through

restriction practice, in clear contradiction to precedential holdings, while usurping the jurisdiction of the Board of Patent Appeals and Interferences and the Federal Circuit to review the policy of § 803.04 and the Office's application thereof. *Appln. of Weber*, 580 F.2d 455, 458 ("[i]f, however, a single claim is required to be divided up. . . that claim would never be considered on its merits."); *Appln. of Haas*, 486 F.2d 1053, 1056 ("[t]he withdrawal from any further consideration was instead classified as a 'refusal. . . to act on claims' in keeping with a restriction requirement and deemed a purely administrative matter.") That is improper. *Appln. of Weber,* 580 F.2d 455, 458 (restriction practice "does not, however, provide a basis for an examiner acting under the authority of the Commissioner to Reject a particular Claim [sic] on that same basis.")

In the present case, the restriction requirement amounted to a rejection of the claimed invention. It is indisputable that the restriction requirement destroyed the whole of the invention and the examination of its parts can never amount to examination of the whole. Moreover, it is well-established that restriction practice is only permissible between claims directed to different inventions and not within a single claim or group of claims directed to a single invention. As such, the restriction requirement imposed on Applicants' microarray amounted to a rejection of that invention.

D. There Is No Undue Burden on the Office to Examine the Claimed Invention

The Office wishes to sustain the restriction requirement of claims 8-10 because it would allegedly present an "undue burden to perform a search on every combination of 100 nucleic acid molecules selected from a set of 400 molecules." Restriction Requirement at page 2. This position cannot prevail.

The rationale above is, first, inaccurate. Although Applicants acknowledge that thousands upon thousands of different collections and combinations are envisioned in the claimed microarray, Applicants submit that the Office should direct its attention to determining

novelty of all of the individual nucleic acid sequences set forth in the group of sequences of claim 8.[5] That is because if they are all novel, all combinations contemplated within the scope of the claimed invention would be novel. If any one of them is found not novel, Applicants would be required to reexamine the parameters of the claimed invention and present either arguments or amendments to comply with the requirements for patentability, as they would in any other invention examined by the Office.

To that end, Applicants here have already provided assistance to the Office by submitting a CD-ROM that includes the results of the search performed by Applicants prior to adding claims 8-10 to the Application. Applicants explained that

> [b]ased on analysis of the BLASTN output the original set of 2921 sequences was reduced by selecting only those sequences which were greater than 400 nucleotides and (a) had no matches to any public sequence in the queried database, or (b) matched for the top hit (best E value) . . . to a public sequence in the queried database in only a single high scoring pair . . . of less than 100 nucleotides where the match had an expectation (E value) greater than $1E^{-3}$.

Preliminary Amendment at page 15. Although the First Decision appears to make jest of Applicants' submission,[6] the reality is that the results of Applicants' search greatly reduces the amount of time and effort required by the Office to determine the novelty of these nucleic acid sequences. In particular, the First Office Action states that "[f]or each claimed SEQ ID Number, the Office must perform a sequence search, for each SEQ ID Number, on a commercial database (which included multiple databases), PTO in-house database, and the issued patent database." First Office Action at page 2. Under this rationale, Applicants have already certified their results for one-third of the work the Office asserts it would need to do.

---

[5] The Second Decision argues that "[b]ecause the Office does not have the resources for such an undertaking, USPTO policy is to require applicants to elect a single combination for examination, as set forth in MPEP 803.04." Second Decision at page 2. That requirement is inconsistent with governing authority, as previously discussed.

[6] The First Decision states "Applicants magnanimously offer to assist in searching the prior art for the sequences listed in the claims, and point out that the USPTO is considering using prior art searches performed by commercial search authorities in the future." First Decision at page 2.

Furthermore, Applicants have submitted evidence that a search of 497 nucleic acid sequences would not pose an ***undue*** burden on the Office, but rather a ***reasonable*** one. In response to the Restriction Requirement, Applicants pointed out that

> It took applicants' representative less than 10 minutes to set up the BLASTN search. After the BLASTN search was completed, Applicants' representative spent approximately 2 hours examining and parsing the BLASTN output with the purpose of selecting those sequences which either had no matches to any sequence in the queried database or which fulfilled other criteria. In this case, the Examiner is being asked to examine only 498 [sic – 497] sequences, rather than approximately 3000. Applicants representative submits that even if the Examiner were to search a larger database than the one described above, or use a different search algorithm, the time for running a program is not synonymous with the actual time that an Examiner would have to spend on a search. To further avoid any undue search burden the PTO is encouraged to refer to the preliminary amendment of October 10, 2000 in which applicant [sic] submitted a copy of the BLASTIN output on CD-ROM.

Response to Restriction Requirement at page 2. To date, the Office has submitted no evidence to contradict Applicants assessment of the ***actual*** time and effort required by the Office to perform a similar search.

In any event, the burden issue is a red herring. For reasons already explained, the Office is obligated to examine Applicants' claimed invention. Restriction between claims may be proper, but restriction within claims is not.

> As a general proposition, an applicant has a right to have Each claim examined on the merits. If an applicant submits a number of claims, it may well be that pursuant to a proper restriction requirement, those claims will be dispersed to a number of applications. . . If, however, a single claim is required to be divided up and presented in several applications, that claim would never be considered on its merits. The totality of the resulting fragmentary claims would not necessarily be the equivalent of the original claim. *Appln. of Weber,* 580 F.2d 455, 458.

And a refusal to examine the invention based on such a restriction constitutes an improper rejection.

> We hold, therefore, that the Board of Appeals has jurisdiction under 35 U.S. C. § § 7 and 134 to review an adverse decision of an examiner when that decision, although designated a 'withdrawal' of a claim from further consideration, is in

> fact a rejection of that claim on the ground that it encompasses independent and distinct inventions. *Appln. of Haas,* 486 F.2d 1053, 1056

Yet, the Office chooses to disregard both the law and Applicants' prior search submitted on CD-ROM. The evidence offered by Applicants supports the position that a search of the nucleic acid sequences listed in claim 8 would not pose an "undue burden". The only position the Office advances in response is that "[a]t the present time, all prior art searches for U.S. patent applications are performed 'in house.' " First Decision at page 2. However, this position is not sustainable. As Applicants have previously pointed out,

> [f]or example, the Office currently provides for an applicant to submit prior art references and certify that he has made or caused to be made a careful and thorough search of the prior art in situations where a Petition to Make Special under 37 C.F.R. § 1.102 has been granted. *See* MPEP § 708.02. Moreover, the Office has acknowledged that '[a]pplicants are generally in the best position to identify the most pertinent prior art related to their invention(s).'

Request at page 8.[7]

When the evidence is considered as a whole, it becomes clear that the only undue burden on the Office regarding the examination of multiple nucleic acid sequences is the burden the Office imposes on itself. Several solutions have been offered on multiple occasions to ease the search and examination burden of the Office. On every occasion, the Office resists these changes. For example, because the issued-patent database is available to interested members of the community, it is not beyond reason to suggest that applicants could perform this portion of the search and certify the results, as well as submit their own searches of publicly available databases. The Office may even easily develop guidelines and requirements that these searches must meet before the Office will consider the results. This solution was offered by several members of the patent community at the Public Hearings on Patenting of Nucleic Acid

---

[7] *See also Four-Tracks Patent Examination Process*, Productivity, Pendency 2 at page 9 ("Many times, applicants possess the expertise to recognize and identify the most pertinent prior art patents and publications related to their invention(s). Review of the ISSR will result in applicant identification of significant patentability issues related to novelty and obviousness before an examiner even begins, thus enabling examiners to better spend their time on the analysis of patentability that is critical to patent quality.")

Sequences ("Hearings") on April 16 and April 23, 1996 (attached hereto as Appendix C). Abuse of this proposed procedure is presently guarded against by an applicant's duty of candor under 37 C.F.R. § 1.56 and by publication of patent applications after 18 months.

Numerous other solutions have been offered by the patent community, both in the Hearings and in preparation of the 21st Century Strategic Plan.[8] Several examples from the Hearings include (1) charging additional fees for the examination of inventions that include sequences in excess of a prescribed amount; (2) eliminating excess sequences from databases used by the Office to avoid redundancy; (3) outsourcing to private, specialized search firms; (4) developing software and/or searching methods for the Office and examiner training; (5) using BLAST searches, such as the one performed by Applicants, and other filtering systems to narrow the number of sequences requiring further search and examination; and (6) two-part examination. Furthermore, adoption of any of one or more of these suggestions would be consummate in scope with the goals expressed in the 21st Century Strategic Plan.

All of these options were originally offered or suggested in the Hearings conducted in 1996. The Office could have taken steps to implement them then and, because of the advancement of technology, could even more easily and cheaply implement them now. However, in 1996, the Office seized on the ability to use a resource already available to them – restriction practice – in spite of warnings from the patent community that this practice was already overzealously applied. Today, the Office still clings to this archaic "quick fix" while simultaneously claiming it would be an "***undue*** burden" to search more than one sequence. This burden is unquestionably self-imposed.

In the present case, Applicants have demonstrated that the current restriction practice, embodied in M.P.E.P. § 803.04, applied to applications containing nucleic acid and amino acid

---

[8] *See,* for example, the quarterly Biotechnology/Chemical/Pharmaceutical Customer Partnership Conferences, designed and developed to be a forum to share ideas, experiences and insights between individual users and the Office.

sequences is improperly imposed to restrict within a single invention and to avoid examination of the invention claimed. Moreover, the Office cannot sustain its position that examination of more than one sequence in an application poses an undue burden. Rather, it is the Office itself that is guilty of imposing an undue burden on Applicants by taking away their opportunity to have what they regard as their invention examined in accordance with the law.

Conclusion

In view of the foregoing, it is respectfully requested that the Board of Patent Appeals and Interferences reverse the Rejections and that the subject application be allowed forthwith.

Respectfully submitted,

Thomas E. Kelley

Thomas E. Kelley (Reg. No. 29, 938)
by David R. Marsh (Reg. No. 41,408)
Holly Logue Prutz (Reg. No. 47,755)

by David R. Marsh

Date: June 30, 2003

MONSANTO COMPANY
62 Maritime Drive
Mystic, CT 06355-1958
860-572-5274 telephone
860-572-5280 facsimile

ARNOLD & PORTER
The Thurman Arnold Building
555 Twelfth Street, N.W.
Washington, D.C. 20004-1206
202.942.5000 telephone
202.942.5999 facsimile

APPENDIX A

8. A microarray comprising a substrate with a surface comprising 10^3 nucleic acid molecules or more where at least 10% of the nucleic acid molecules are comprised of different sequences and at least about 250 nucleotide residues and complementary to a molecule comprising a sequence selected from the group consisting of SEQ ID NO: 5776, SEQ ID NO: 5781, SEQ ID NO: 5782, SEQ ID NO: 5783, SEQ ID NO: 5785, SEQ ID NO: 5787, SEQ ID NO: 5800, SEQ ID NO: 5804, SEQ ID NO: 5815, SEQ ID NO: 5818, SEQ ID NO: 5821, SEQ ID NO: 5823, SEQ ID NO: 5828, SEQ ID NO: 5830, SEQ ID NO: 5832, SEQ ID NO: 5836, SEQ ID NO: 5838, SEQ ID NO: 5840, SEQ ID NO: 5845, SEQ ID NO: 5849, SEQ ID NO: 5850, SEQ ID NO: 5851, SEQ ID NO: 5856, SEQ ID NO: 5859, SEQ ID NO: 5863, SEQ ID NO: 5868, SEQ ID NO: 5871, SEQ ID NO: 5874, SEQ ID NO: 5875, SEQ ID NO: 5877, SEQ ID NO: 5893, SEQ ID NO: 5896, SEQ ID NO: 5901, SEQ ID NO: 5908, SEQ ID NO: 5909, SEQ ID NO: 5920, SEQ ID NO: 5922, SEQ ID NO: 5926, SEQ ID NO: 5928, SEQ ID NO: 5929, SEQ ID NO: 5931, SEQ ID NO: 5936, SEQ ID NO: 5937, SEQ ID NO: 5939, SEQ ID NO: 5941, SEQ ID NO: 5944, SEQ ID NO: 5945, SEQ ID NO: 5950, SEQ ID NO: 5955, SEQ ID NO: 5960, SEQ ID NO: 5961, SEQ ID NO: 5963, SEQ ID NO: 5964, SEQ ID NO: 5968, SEQ ID NO: 5973, SEQ ID NO: 5974, SEQ ID NO: 5991, SEQ ID NO: 5994, SEQ ID NO: 5999, SEQ ID NO: 6000, SEQ ID NO: 6001, SEQ ID NO: 6005, SEQ ID NO: 6006, SEQ ID NO: 6007, SEQ ID NO: 6011, SEQ ID NO: 6017, SEQ ID NO: 6018, SEQ ID NO: 6022, SEQ ID NO: 6023, SEQ ID NO: 6026, SEQ ID NO: 6030, SEQ ID NO: 6033, SEQ ID NO: 6042, SEQ ID NO: 6046, SEQ ID NO: 6059, SEQ ID NO: 6063, SEQ ID NO: 6065, SEQ ID NO: 6066, SEQ ID NO: 6089, SEQ ID NO: 6091, SEQ ID NO: 6098, SEQ ID NO: 6106, SEQ ID NO: 6107, SEQ ID NO: 6110, SEQ ID NO: 6117, SEQ ID NO: 6121, SEQ ID NO: 6124, SEQ ID NO: 6131, SEQ ID NO: 6137, SEQ ID NO: 6141, SEQ ID NO: 6144, SEQ ID NO: 6145, SEQ ID NO: 6147, SEQ ID NO: 6154, SEQ ID NO: 6167, SEQ ID NO: 6168, SEQ ID NO: 6170, SEQ ID NO: 6173, SEQ ID NO: 6178, and SEQ ID NO: 6181.

9. A microarray according to claim 8 where at least 75% of the nucleic acid molecules are comprised of different sequences and at least about 250 nucleotide residues and complementary to a molecule comprising a sequence selected from said group.

10. A microarray according to claim 8 where at least 95% of the nucleic acid molecules are comprised of different sequences and at least about 250 nucleotide residues and complementary to a molecule comprising a sequence selected from said group.

Appendix B

Claims as originally filed in the Preliminary Amendment of October 10, 2000

8. A microarray having a substrate with a surface comprising 10^3 nucleic acid molecules or more where at least 10% of the nucleic acid molecules are different and at least about 250 nucleotide residues and complementary to a molecule having a sequence selected from the group consisting of SEQ ID NO: 5776 and SEQ ID NO: 5781 and SEQ ID NO: 5782 and SEQ ID NO: 5783 and SEQ ID NO: 5786 and SEQ ID NO: 5787 and SEQ ID NO: 5800 and SEQ ID NO: 5815 and SEQ ID NO: 5818 and SEQ ID NO: 5821 and SEQ ID NO: 5823 and SEQ ID NO: 5828 and SEQ ID NO: 5830 and SEQ ID NO: 5836 and SEQ ID NO: 5838 and SEQ ID NO: 5840 and SEQ ID NO: 5845 and SEQ ID NO: 5849 and SEQ ID NO: 5850 and SEQ ID NO: 5851 and SEQ ID NO: 5859 and SEQ ID NO: 5863 and SEQ ID NO: 5868 and SEQ ID NO: 5874 and SEQ ID NO: 5875 and SEQ ID NO: 5877 and SEQ ID NO: 5893 and SEQ ID NO: 5896 and SEQ ID NO: 5901 and SEQ ID NO: 5909 and SEQ ID NO: 5922 and SEQ ID NO: 5926 and SEQ ID NO: 5928 and SEQ ID NO: 5931 and SEQ ID NO: 5936 and SEQ ID NO: 5937 and SEQ ID NO: 5939 and SEQ ID NO: 5941 and SEQ ID NO: 5950 and SEQ ID NO: 5955 and SEQ ID NO: 5956 and SEQ ID NO: 5963 and SEQ ID NO: 5973 and SEQ ID NO: 5974 and SEQ ID NO: 5991 and SEQ ID NO: 5994 and SEQ ID NO: 5999 and SEQ ID NO: 6000 and SEQ ID NO: 6001 and SEQ ID NO: 6005 and SEQ ID NO: 6006 and SEQ ID NO: 6007 and SEQ ID NO: 6011 and SEQ ID NO: 6017 and SEQ ID NO: 6022 and SEQ ID NO: 6023 and SEQ ID NO: 6030 and SEQ ID NO: 6033 and SEQ ID NO: 6059 and SEQ ID NO: 6065 and SEQ ID NO: 6089 and SEQ ID NO: 6091 and SEQ ID NO: 6106 and SEQ ID NO: 6107 and SEQ ID NO: 6110 and SEQ ID NO: 6117 and SEQ ID NO: 6121 and SEQ ID NO: 6124 and SEQ ID NO: 6137 and SEQ ID NO: 6154 and SEQ ID NO: 6167 and SEQ ID NO: 6168 and SEQ ID NO: 6170 and SEQ ID NO: 6173 and SEQ ID NO: 6178 and SEQ ID NO: 6181 and SEQ ID NO: 6188 and SEQ ID NO: 6195 and SEQ ID NO: 6196 and SEQ ID NO: 6205 and SEQ ID NO: 6211 and SEQ ID NO: 6212 and SEQ ID NO: 6214 and SEQ ID NO: 6234 and SEQ ID NO: 6241 and SEQ ID NO: 6245 and SEQ ID NO: 6251 and SEQ ID NO: 6256 and SEQ ID NO: 6261 and SEQ ID NO: 6270 and SEQ ID NO: 6272 and SEQ ID NO:

6278 and SEQ ID NO: 6283 and SEQ ID NO: 6286 and SEQ ID NO: 6288 and SEQ ID NO: 6289 and SEQ ID NO: 6291 and SEQ ID NO: 6292 and SEQ ID NO: 6293 and SEQ ID NO: 6295 and SEQ ID NO: 6302 and SEQ ID NO: 6309 and SEQ ID NO: 6316 and SEQ ID NO: 6321 and SEQ ID NO: 6324 and SEQ ID NO: 6332 and SEQ ID NO: 6337 and SEQ ID NO: 6347 and SEQ ID NO: 6348 and SEQ ID NO: 6352 and SEQ ID NO: 6355 and SEQ ID NO: 6358 and SEQ ID NO: 6361 and SEQ ID NO: 6363 and SEQ ID NO: 6367 and SEQ ID NO: 6369 and SEQ ID NO: 6371 and SEQ ID NO: 6375 and SEQ ID NO: 6377 and SEQ ID NO: 6385 and SEQ ID NO: 6405 and SEQ ID NO: 6416 and SEQ ID NO: 6418 and SEQ ID NO: 6425 and SEQ ID NO: 6433 and SEQ ID NO: 6445 and SEQ ID NO: 6451 and SEQ ID NO: 6454 and SEQ ID NO: 6459 and SEQ ID NO: 6462 and SEQ ID NO: 6464 and SEQ ID NO: 6465 and SEQ ID NO: 6466 and SEQ ID NO: 6476 and SEQ ID NO: 6479 and SEQ ID NO: 6481 and SEQ ID NO: 6482 and SEQ ID NO: 6483 and SEQ ID NO: 6489 and SEQ ID NO: 6497 and SEQ ID NO: 6502 and SEQ ID NO: 6514 and SEQ ID NO: 6521 and SEQ ID NO: 6529 and SEQ ID NO: 6530 and SEQ ID NO: 6535 and SEQ ID NO: 6540 and SEQ ID NO: 6553 and SEQ ID NO: 6556 and SEQ ID NO: 6564 and SEQ ID NO: 6565 and SEQ ID NO: 6571 and SEQ ID NO: 6598 and SEQ ID NO: 6603 and SEQ ID NO: 6604 and SEQ ID NO: 6606 and SEQ ID NO: 6611 and SEQ ID NO: 6613 and SEQ ID NO: 6620 and SEQ ID NO: 6628 and SEQ ID NO: 6633 and SEQ ID NO: 6639 and SEQ ID NO: 6640 and SEQ ID NO: 6643 and SEQ ID NO: 6651 and SEQ ID NO: 6652 and SEQ ID NO: 6654 and SEQ ID NO: 6657 and SEQ ID NO: 6659 and SEQ ID NO: 6661 and SEQ ID NO: 6664 and SEQ ID NO: 6668 and SEQ ID NO: 6670 and SEQ ID NO: 6671 and SEQ ID NO: 6674 and SEQ ID NO: 6680 and SEQ ID NO: 6685 and SEQ ID NO: 6686 and SEQ ID NO: 6688 and SEQ ID NO: 6691 and SEQ ID NO: 6692 and SEQ ID NO: 6696 and SEQ ID NO: 6697 and SEQ ID NO: 6702 and SEQ ID NO: 6704 and SEQ ID NO: 6711 and SEQ ID NO: 6716 and SEQ ID NO: 6724 and SEQ ID NO: 6726 and SEQ ID NO: 6728 and SEQ ID NO: 6739 and SEQ ID NO: 6742 and SEQ ID NO: 6744 and SEQ ID NO: 6747 and SEQ ID NO: 6764 and SEQ ID NO: 6766 and SEQ ID NO: 6770 and SEQ ID NO: 6772 and SEQ ID NO: 6773 and SEQ ID NO: 6778 and SEQ ID NO: 6781 and SEQ ID NO: 6782 and SEQ ID NO: 6790 and SEQ ID NO: 6792 and SEQ ID NO: 6841 and SEQ ID NO: 6870 and SEQ ID NO: 6871 and SEQ ID NO:

6886 and SEQ ID NO: 6894 and SEQ ID NO: 6904 and SEQ ID NO: 6906 and SEQ ID NO: 6923 and SEQ ID NO: 6931 and SEQ ID NO: 6934 and SEQ ID NO: 6938 and SEQ ID NO: 6939 and SEQ ID NO: 6949 and SEQ ID NO: 6953 and SEQ ID NO: 6961 and SEQ ID NO: 6965 and SEQ ID NO: 6978 and SEQ ID NO: 6991 and SEQ ID NO: 7000 and SEQ ID NO: 7002 and SEQ ID NO: 7036 and SEQ ID NO: 7050 and SEQ ID NO: 7053 and SEQ ID NO: 7062 and SEQ ID NO: 7070 and SEQ ID NO: 7078 and SEQ ID NO: 7083 and SEQ ID NO: 7096 and SEQ ID NO: 7100 and SEQ ID NO: 7106 and SEQ ID NO: 7107 and SEQ ID NO: 7134 and SEQ ID NO: 7144 and SEQ ID NO: 7161 and SEQ ID NO: 7164 and SEQ ID NO: 7178 and SEQ ID NO: 7206 and SEQ ID NO: 7220 and SEQ ID NO: 7221 and SEQ ID NO: 7224 and SEQ ID NO: 7228 and SEQ ID NO: 7229 and SEQ ID NO: 7241 and SEQ ID NO: 7246 and SEQ ID NO: 7248 and SEQ ID NO: 7249 and SEQ ID NO: 7254 and SEQ ID NO: 7261 and SEQ ID NO: 7267 and SEQ ID NO: 7277 and SEQ ID NO: 7332 and SEQ ID NO: 7338 and SEQ ID NO: 7351 and SEQ ID NO: 7357 and SEQ ID NO: 7369 and SEQ ID NO: 7378 and SEQ ID NO: 7380 and SEQ ID NO: 7384 and SEQ ID NO: 7386 and SEQ ID NO: 7395 and SEQ ID NO: 7438 and SEQ ID NO: 7444 and SEQ ID NO: 7447 and SEQ ID NO: 7461 and SEQ ID NO: 7463 and SEQ ID NO: 7465 and SEQ ID NO: 7468 and SEQ ID NO: 7478 and SEQ ID NO: 7490 and SEQ ID NO: 7491 and SEQ ID NO: 7492 and SEQ ID NO: 7509 and SEQ ID NO: 7514 and SEQ ID NO: 7536 and SEQ ID NO: 7559 and SEQ ID NO: 7563 and SEQ ID NO: 7565 and SEQ ID NO: 7572 and SEQ ID NO: 7576 and SEQ ID NO: 7590 and SEQ ID NO: 7592 and SEQ ID NO: 7594 and SEQ ID NO: 7596 and SEQ ID NO: 7602 and SEQ ID NO: 7612 and SEQ ID NO: 7617 and SEQ ID NO: 7618 and SEQ ID NO: 7622 and SEQ ID NO: 7624 and SEQ ID NO: 7627 and SEQ ID NO: 7631 and SEQ ID NO: 7632 and SEQ ID NO: 7644 and SEQ ID NO: 7646 and SEQ ID NO: 7655 and SEQ ID NO: 7657 and SEQ ID NO: 7659 and SEQ ID NO: 7677 and SEQ ID NO: 7682 and SEQ ID NO: 7685 and SEQ ID NO: 7690 and SEQ ID NO: 7694 and SEQ ID NO: 7697 and SEQ ID NO: 7701 and SEQ ID NO: 7702 and SEQ ID NO: 7703 and SEQ ID NO: 7707 and SEQ ID NO: 7708 and SEQ ID NO: 7711 and SEQ ID NO: 7715 and SEQ ID NO: 7718 and SEQ ID NO: 7721 and SEQ ID NO: 7723 and SEQ ID NO: 7724 and SEQ ID NO: 7729 and SEQ ID NO: 7736 and SEQ ID NO: 7745 and SEQ ID NO: 7749 and SEQ ID NO: 7757 and SEQ ID NO:

7761 and SEQ ID NO: 7770 and SEQ ID NO: 7772 and SEQ ID NO: 7777 and SEQ ID NO: 7782 and SEQ ID NO: 7791 and SEQ ID NO: 7795 and SEQ ID NO: 7796 and SEQ ID NO: 7800 and SEQ ID NO: 7801 and SEQ ID NO: 7802 and SEQ ID NO: 7808 and SEQ ID NO: 7809 and SEQ ID NO: 7813 and SEQ ID NO: 7818 and SEQ ID NO: 7823 and SEQ ID NO: 7825 and SEQ ID NO: 7826 and SEQ ID NO: 7827 and SEQ ID NO: 7831 and SEQ ID NO: 7837 and SEQ ID NO: 7844 and SEQ ID NO: 7845 and SEQ ID NO: 7847 and SEQ ID NO: 7849 and SEQ ID NO: 7852 and SEQ ID NO: 7856 and SEQ ID NO: 7860 and SEQ ID NO: 7862 and SEQ ID NO: 7863 and SEQ ID NO: 7865 and SEQ ID NO: 7867 and SEQ ID NO: 7870 and SEQ ID NO: 7872 and SEQ ID NO: 7880 and SEQ ID NO: 7883 and SEQ ID NO: 7884 and SEQ ID NO: 7885 and SEQ ID NO: 7886 and SEQ ID NO: 7888 and SEQ ID NO: 7889 and SEQ ID NO: 7894 and SEQ ID NO: 7895 and SEQ ID NO: 7896 and SEQ ID NO: 7901 and SEQ ID NO: 7905 and SEQ ID NO: 7908 and SEQ ID NO: 7918 and SEQ ID NO: 7928 and SEQ ID NO: 7931 and SEQ ID NO: 7937 and SEQ ID NO: 7947 and SEQ ID NO: 7952 and SEQ ID NO: 7959 and SEQ ID NO: 7962 and SEQ ID NO: 7975 and SEQ ID NO: 7987 and SEQ ID NO: 7989 and SEQ ID NO: 7992 and SEQ ID NO: 7997 and SEQ ID NO: 7998 and SEQ ID NO: 8002 and SEQ ID NO: 8006 and SEQ ID NO: 8007 and SEQ ID NO: 8008 and SEQ ID NO: 8010 and SEQ ID NO: 8015 and SEQ ID NO: 8017 and SEQ ID NO: 8018 and SEQ ID NO: 8020 and SEQ ID NO: 8022 and SEQ ID NO: 8024 and SEQ ID NO: 8026 and SEQ ID NO: 8043 and SEQ ID NO: 8048 and SEQ ID NO: 8053 and SEQ ID NO: 8063 and SEQ ID NO: 8064 and SEQ ID NO: 8066 and SEQ ID NO: 8095 and SEQ ID NO: 8098 and SEQ ID NO: 8099 and SEQ ID NO: 8103 and SEQ ID NO: 8126 and SEQ ID NO: 8145 and SEQ ID NO: 8153 and SEQ ID NO: 8155 and SEQ ID NO: 8157 and SEQ ID NO: 8161 and SEQ ID NO: 8171 and SEQ ID NO: 8175 and SEQ ID NO: 8180 and SEQ ID NO: 8184 and SEQ ID NO: 8192 and SEQ ID NO: 8198 and SEQ ID NO: 8199 and SEQ ID NO: 8201 and SEQ ID NO: 8212 and SEQ ID NO: 8216 and SEQ ID NO: 8222 and SEQ ID NO: 8224 and SEQ ID NO: 8225 and SEQ ID NO: 8228 and SEQ ID NO: 8238 and SEQ ID NO: 8242 and SEQ ID NO: 8256 and SEQ ID NO: 8259 and SEQ ID NO: 8274 and SEQ ID NO: 8277 and SEQ ID NO: 8278 and SEQ ID NO: 8289 and SEQ ID NO: 8292 and SEQ ID NO: 8297 and SEQ ID NO: 8301 and SEQ ID NO: 8317 and SEQ ID NO: 8329 and SEQ ID NO:

8333 and SEQ ID NO: 8334 and SEQ ID NO: 8335 and SEQ ID NO: 8351 and SEQ ID NO: 8355 and SEQ ID NO: 8374 and SEQ ID NO: 8377 and SEQ ID NO: 8379 and SEQ ID NO: 8383 and SEQ ID NO: 8394 and SEQ ID NO: 8406 and SEQ ID NO: 8447 and SEQ ID NO: 8452 and SEQ ID NO: 8455 and SEQ ID NO: 8458 and SEQ ID NO: 8465 and SEQ ID NO: 8468 and SEQ ID NO: 8472 and SEQ ID NO: 8479 and SEQ ID NO: 8482 and SEQ ID NO: 8487 and SEQ ID NO: 8492 and SEQ ID NO: 8500 and SEQ ID NO: 8503 and SEQ ID NO: 8507 and SEQ ID NO: 8511 and SEQ ID NO: 8512 and SEQ ID NO: 8517 and SEQ ID NO: 8518 and SEQ ID NO: 8529 and SEQ ID NO: 8530 and SEQ ID NO: 8538 and SEQ ID NO: 8542 and SEQ ID NO: 8553 and SEQ ID NO: 8554 and SEQ ID NO: 8556 and SEQ ID NO: 8560 and SEQ ID NO: 8568 and SEQ ID NO: 8569 and SEQ ID NO: 8578 and SEQ ID NO: 8579 and SEQ ID NO: 8580 and SEQ ID NO: 8583 and SEQ ID NO: 8584 and SEQ ID NO: 8585 and SEQ ID NO: 8587 and SEQ ID NO: 8590 and SEQ ID NO: 8601 and SEQ ID NO: 8607 and SEQ ID NO: 8611 and SEQ ID NO: 8616 and SEQ ID NO: 8624 and SEQ ID NO: 8625 and SEQ ID NO: 8631 and SEQ ID NO: 8632 and SEQ ID NO: 8639 and SEQ ID NO: 8644 and SEQ ID NO: 8665.

9. A microarray according to claim 8 where at least 75% of the nucleic acid molecules are different and at least about 250 nucleotide residues and complementary to a molecule having a sequence selected from said group.

10. A microarray according to claim 8 where at least 95% of the nucleic acid molecules are different and at least about 250 nucleotide residues and complementary to a molecule having a sequence selected from said group.

---

[9] This C.C.P.A. precedent is binding. *South Corp. v. U.S.*, 690 F.2d 1368, 1369, 215 U.S.P.Q. 657 ("[w]e hold that the holdings of our predecessor courts, the United States Court of Claims and the United States Court of Customs and Patent Appeals, announced by those courts before the close of business September 30, 1982, shall be binding as precedent in this court.")

[10] For example, at a minimum, if Applicants were to file 497 applications, each one directed to a microarray which must include a specific SEQ ID NO, as the Office has done here (*see* Sec, 8.D and footnote 5 *infra*), the cost to Applicants would be $ 372,750 in initial filing fees for the Office alone. Furthermore, even this expenditure would not allow Applicants' to have what they regard as the invention examined.

[11] Applicants stated:

> To the extent that this is not intended to be a further restriction requirement, but merely an election of species for examination, applicant elects the first 100 sequences in the Markush group.

Response to Restriction Requirement at page 3. This response was held by the Office as non-responsive and Applicants were required explicitly recite the 100 nucleic acid sequences selected.

[12] Accordingly, the examined invention no longer resembled Applicants' claimed invention, but rather a microarray that always included SEQ ID NO: 5893 – no matter what other sequences were included in the claim.

[13] The Second Decision argues that "[b]ecause the Office does not have the resources for such an undertaking, USPTO policy is to require applicants to elect a single combination for examination, as set forth in MPEP 803.04." Second Decision at page 2. That requirement is inconsistent with governing authority, as previously discussed.

[14] The First Decision states "Applicants magnanimously offer to assist in searching the prior art for the sequences listed in the claims, and point out that the USPTO is considering using prior art searches performed by commercial search authorities in the future." First Decision at page 2.

[15] *See also Four-Tracks Patent Examination Process*, Productivity, Pendency 2 at page 9 ("Many times, applicants possess the expertise to recognize and identify the most pertinent prior art patents and publications related to their invention(s). Review of the ISSR will result in applicant identification of significant patentability issues related to novelty and obviousness before an examiner even begins, thus enabling examiners to better spend their time on the analysis of patentability that is critical to patent quality.")

[16] footnote name of meetings – Tom will send

HEARING AND REQUEST FOR COMMENTS ON ISSUES RELATING

TO PATENT PROTECTION FOR NUCLEIC ACID SEQUENCES

University of California, San Diego

International Center

April 16, 1996

Reported By: TERESA BURT, CSR NO. 8607

PANEL MEMBERS:

Bruce A. Lehman
Stephen G. Kunin
Edward R. Kazenske
Lawrence J. Goffney
Nancy J. Linck

MR. LEHMAN: My name is Bruce Lehman. I'm the assistant secretary of commerce and commissioner of patents and trademarks.

And joining me today in this hearing are Lawrence J. Goffney, acting deputy district secretary of commerce and acting deputy commissioner of patents and trademarks.

To my immediate right, Edward Kazenske, deputy assistant commissioner for patents.

To my left, Stephen Kunin, deputy assistant commissioner for patent, policy and projects.

Over here on my far right, Nancy Linck, the solicitor, the chief lawyer of the Patent and Trademark Office.

I'd like to first thank the University of California at San Diego for providing this facility for us. This is the second time we've been out here in this campus.

This is a hearing to receive public comment on a serious problem currently facing the Patent and Trademark Office related to patent protection for nucleic acid sequences.

The public was invited to comment on this issue in a notice published on March 12th, 1996, in Volume 61, No. 49 of the Federal Register.

For over a decade, the Patent and Trademark Office has been examining and granting patents to claims

receiting nucleic acid sequences. The scientific and technological advances have permitted the rapid

identification of large numbers of genes or gene fragments. The ease of utilizing automated techniques for sequencing nucleic acid fragments has resulted in the filing of a growing, although still relatively small number, of patent applications, each of which claim thousands of nucleic acid sequences.

Statistics reveal that the number of these applications is growing, and based on the number of organisms in genes still to be discovered, such growth will continue for the near future.

In fiscal year 1991, the scientific and technical information center of our Patent and Trademark Office searched about 4,000 sequences. In fiscal year '95, they searched about 22,000 sequences. Currently we have over 200,000 sequences claimed and at least 70 patent applications awaiting search and examination.

Our estimates show that the search of 100 sequences requires about 15 hours of computer time. But the evaluation of the search results for those 100 sequences requires about 65 hours of examiner time. The PTO currently has two massively parallel processor computers and could run the searches in about two years, when the computers are running 24 hours a day, seven days a week.

To examine this relatively small number of patent applications only with respect to the prior art, however, would require only 9- -- would require over 90 senior-level staff years. Thus, in order to accommodate --

in order to process these applications, the entire staff of the Biotechnology Patent Examining Group 1800 would have to work for more than nine months exclusively on these applications.

These applications present a challenge to us at the PTO, and we need help and suggestions on how we can address the problem. The United States is a leader in the rapidly growing field of biotechnology, which is a growth industry important to the economic health of this country.

The PTO has taken a very active role in

working with its customers to try to simplify policies and procedures in ways that encourage and promote the growth of this industry.

We are committed to improving the responsiveness of the PTO to its customers and to more effectively address the needs of industry. We must find ways to search and examine the pending applications and provide these applicants with appropriate protection -- and I emphasize the word appropriate patent protection -- for the inventions without creating an imbalance in the appropriation of the resources within and among the technologies and the patent-examining groups.

These policies established must permit the timely and thorough examination of all applications which require the same resources for completion.

We are currently working in partnership with the applicants of these applications in order to explore innovative mechanisms to accomplish the required work in

processing them. We appreciate the time each of you has taken to attend this hearing -- all of you in the audience today -- and provide us with input into the solutions to this problem.

A transcript of this hearing will be prepared and will be made available for purchase by the public approximately 10 days after this hearing. Copies of the transcript will also be available for purchase directly from the stenographer. The name of the stenographer service today is Peterson & Associates, and their telephone number is area code (619) 260-1069. That's (619) 260-1069, Peterson & Associates.

We have received 10 written comments and seven requests to appear orally at this hearing. Any persons who wish to speak and who have not previously informed us of their desire are encouraged to add their names to the list located on the table at the rear of this room.

In order to permit all persons requesting to appear orally, including those persons signing up today to present testimony, we would request each speaker to limit their presentation to 15 minutes. Those persons who wish

to provide additional comments should submit their comments in writing to us no later than April 23rd.

The speakers have been listed, and I think we have a list of them back there, don't we?

They've been listed in the order in which the requests were received by us. You may also pick up at the table of the rear of the room copies of the Official

Gazette publication of the notice of these hearings and request for comments on issues related to patent protection for nucleic acid sequences.

When you present your comments, we would appreciate it if you would please give your name and address and tell us whether these comments are your own or whether they are those of a law firm or company that you represent or whether you represent an organization and are presenting comments on their behalf.

Speakers are requested to limit their remarks to the questions presented in the Federal Register publication of March 12th, 1996. And I am told that the first person who is here to testify before us -- where is -- oh, right there -- to testify before us this morning is Alice Martin.

So, Ms. Martin, would you please come forward.

MS. MARTIN: Commissioner, ladies and gentlemen, my name is Alice O. Martin. I represent the law firm of Brinks, Hofer, Gilson & Lione in Chicago.

My address there is the NBC towers. I also believe I am speaking consistent with the position of B.I.O., of which our firm is a member.

As attorneys for the biotechnology industry, we applaud the cry from the patent office for help in solving what we consider a very common problem. We appreciate that the examiners are hacking their way through this jungle of nucleotide and amino acid sequence databases and that they face a seemingly unmanageable task and,

indeed, as the commissioner mentioned, the jungle is growing even as we speak.

Now, some facts affecting this problem are the database -- databases that exist were not specifically designed for the use of obtaining prior art for obviousness

rejections but rather they were designed to accumulate and expand on all information on all sequences that are known.

The tools for the search algorithms to hack through this jungle were not specifically designed to find prior art for obviousness rejections, but rather they were designed to quantify and describe similarities, sometimes based on an evolutionary model. And many of them are designed to ferret out even distant evolutionary relationships.

Another fact addressing this problem is that examiners are currently, as I understand it, not provided with maps or search strategies with which to find their way through this database jungle. This means that they can take false paths, hit dead ends, waste valuable time and become frustrated.

Furthermore, the subjective, somewhat unpredictable paths that are basically unreproducible cause uncertainty among applicants and certainly down the road will cause problems in litigation in the courts.

Now, how can we make manageable, yet legally sound, solutions to this problem? I suggest we tame the jungle and provide maps through which we could move through the jungle. Now, the tools or the search algorithms used

will to some extent depend on the choice of maps.

How can we tame the jungle? We're not aware of any legal requirement that the patent office has to do an exhaustive, brute-force search of everything that exists in the current databases. So let's prune the jungle. Let's not search all databases just because they exist, particularly when many of them are repetitive.

Perhaps your meeting with experts, as I think has already begun, to explain the construction of the different databases would be helpful so that either you can pick the best database for overall patent searches or pick the best for certain sets of claims. Claims to CD&As may not be using the same database as claims to ESTs, for example. And there are specialized databases that have been developed, for example, by the NIH group for these specialized searches. Perhaps working with these groups

already in the field will be helpful.

Now, a more extreme suggestion that could be necessary is to plan an entirely new jungle.

MR. LEHMAN: Can I ask a question?

MS. MARTIN: Certainly.

MR. LEHMAN: You said that the NIH has developed these specialized databases for the searches, but the NIH has withdrawn their patent applications. Are you talking about databases they developed for other purposes or databases --

MS. MARTIN: That is my understanding. And I'm not up on the latest but I certainly --

MR. LEHMAN: -- they developed for other purposes for their own research purposes?

MS. MARTIN: I have read that they were working on specialized databases, and I thought it might be worth exploring, at least whether they have them available or not, what methods they have used to develop those might be very helpful.

As I said, a more extreme solution would be to completely plan a new jungle. In this case we would take the old databases and create a new database specifically designed for the purpose of finding prior art upon which to base obviousness rejections.

This way perhaps a grouping in the database could be such that you don't have to keep redoing the same kinds of searches over and over, and I'll say a little bit later about the maps that could be used to perhaps organize this database.

We also then in the new database could organize it to correspond to different types of claims that are traditionally sent in by applicants. Also it was mentioned, I believe at the April 4th presentation, that there's a problem in determining what is, quote, prior art because of the difficulty in determining what dates various sequences are entered. That information is in the database but is evidently difficult to retrieve, so art can be looked at but really isn't prior art in the current methods.

A new database may also help the problem

evidently of finding extension numbers and then trying to look in the references that are cited and identify which sequence is really referred to in a particular reference.

And finally, one of the reasons I think the patent office gave for using its present databases is that there are references that are retrieved along with the sequences.

That may not be something that needs to be done for each sequence and, again, can be considered in a new database.

I think everyone agrees that redundant sequences should be removed, and this would be solved either by pruning the databases or developing a new one. Not everyone agrees that sequences should be trimmed first before a search of areas that can give hits that, as some of the developers of the search algorithms say, are of dubious biological significance. And here again it might be helpful to reevaluate the tools that are being used for searches.

My understanding is that the PTO has selected a search algorithm because of increased sensitivity. This is puzzling to us because for increased sensitivity, you are picking up more distantly related sequences, when for obviousness rejection, what we really have to focus on is structurally similar sequences. The more distant the relationships, the more likely there are to be structural dissimilarities, making it more difficult to prove obviousness because, number one, the structures are very

different and, number two, to find art that would suggest or motivate all of the different changes between that distant-claimed and the period-claimed sequence would be very, very rare. So it seems to me that may be defeating the purpose.

The patent office search goals appear to be somewhat different from those we use in the scientific community, and I realize that many of the examiners themselves were established scientists and may have to remove their scientific hat and focus not on the biological relationships but what is structurally similar in terms of

the legal sense. And here again I have a seminar from technical experts who have developed the different algorithms. Could be very, very helpful.

Also, remember that any of these algorithms will find matches because they're really just looking at the sequences as letters and basically searching through to find a probabilistic model that says this sequence is more similar than would be expected by chance. So in and of itself, there is usually no legal or biological significance.

And finally, I suggest we develop maps or search strategies rather than jumping into the jungle first and emerging overburdened and overwhelmed with lots of sequences. We need a strategy because searches in the mechanical arts, if you find an application that has a carburetor and a machine, you don't generally look at all the art in the world that has a carburetor.

13

Similarly in the chemical arts, if you find a composition that has a benzene ring, you don't usually look at all compositions that have a benzene ring.

One option is to use key words first before going into a strategy of searching the database. Key words can include species, library, protein or function as related in the claims and the specification. I think considering the claim as a whole, as the courts and the board of appeal have instructed us, may suggest that some sequences don't even have to be searched.

We could group sequences, and the applicants here should be helpful in this regard in terms of length, core chemical structure, again to relate to the chemical practice art. If we do these maps, this will provide focus using legal standards that already exist. There's nothing new, I think, in the obviousness of DNA sequences.

As you all know, this will be less open to criticism and less open to challenges to validity of the patents than using something such as "Let's look at the first 45 best hits that we get out of the search." That seems to me very arbitrary and I think will open the patent up to much criticism in the future.

Plus, by using that kind of ab priori and

arbitrary criterium, you may have too many sequences examined for some and too little for others. As we saw in the U.S. Supreme Court decision in Deere, there are two criteria that must establish a prima facia case of obviousness. One, is the claim structure different. And

14

two, are the differences suggested by the art.

Both the courts and the board have said DNA is a chemical. Proteins are chemicals, so we must follow the chemical practice, which is the same as iterated in Deere: is the composition structurally similar? And if so, similarity alone will not be a bar unless there are secondary references motivating those changes.

For example, a nucleotide change in two positions must have some reference or some motivation in the art to make those particular differences. As we see it, the existence of these massive databases pose a big risk for hindsight and obviousness to try creeping into obviousness rejections. Looking at the time the invention was made, add a database that exists in the world, the inventor is going to be faced with infinite number of possibilities, and it would be unpredictable to merge or to change sequences without some motivation in the art. So we ask that the examiners take caution in selecting these merely by hindsight.

Some other things that might be kept in mind in planning our sequences searches are that the courts have already said that defining a sequence and a method of making it will not support an obviousness rejection. One species will not necessarily make obvious a similar sequence in a second species.

Here our big problem, I think, boils down to what is structural similarity in terms of DNA and amino acid sequences that can be applied then to the chemical

15

practice standards. We need to decide what is a structural similarity to focus the search rather than getting all sequences that may have some limited degree of similarity. In other words, why look for all sequences because they are somewhat similar if you're not able going to be able to use them in an obvious rejection anyway.

Dealing with some of the smaller sequences that we've mentioned that are presented in large numbers, a portion of the sequence is not likely to make a larger sequence obvious because there is no way of predicting that larger sequence, and the courts are certainly going in that direction. If you have a larger sequence, why look for small fragments when there is no teaching of boundaries of the fragment.

I think here, too, that applicants may have to take a role in being more second specific in some terms that are now fairly loosely defined such as homologues, analogues, mutants, variants and fragments. These are not mutually exclusive terms. They need to be defined.

Now, I know none of us like a piecemeal prosecution. However, there may have to be some negotiation on a 112 section first before a search strategy is planned.

In summary, rather than trying to speed up previous ways to do these searches, particularly before encountering these large numbers or large size sequences, we recommend that we reevaluate whether we shouldn't tame the jungle by finding more suitable databases or groups of

databases for the legal search and by developing maps based on obvious criteria already known to the courts and the board so that we can focus the search and not waste valuable time looking at dead ends. Thank you.

MR. LEHMAN: Thank you, Ms. Martin.

Are there any questions from other members of the panel?

Thank you very much.

Next I'd like to call on John Burke.

MR. BURKE: Good morning. My name is John Burke. I'm here from Maas Parr Computer Corporation. Thank you for having us. I'm going to offer a more technical presentation that won't have much legalities in it.

Maas Parr is a database processing company and in support of bioinformatics of a particular --

THE REPORTER: I need you to speak up a little bit more. Okay?

MR. BURKE: Okay.

Maas Parr Computer is a computer database processing company.

MR. LEHMAN: Could maybe put this closer to your mouth. Okay?

MR. BURKE: Okay. Thank you.

And we also do bioinformatics. We are a complete computer system with fast processing as well as the infrastructure that's required by central processing. Our internal communication is 28 megabytes per second, and

more importantly we have audio capabilities which are needed at the U.S. PTO. We have up to 240 megabytes transferred from the disk to the computer core which will be very important doing the number of searches that you have to perform.

The advantages to U.S. PTO are a quick access to the database, very flexible access, which is very important of your current model of operation. Because you have many independent investigators working on different patent applications, it's not possible to order the database searches. And with many architectures, it's necessary to preload the database. And therefore to optimize your search time, you would have to perform the searches in a certain order on certain databases. And with our systems, this isn't necessary.

As well, the entire range of bioinformatics applications are accelerated, not just the database searching. As you mentioned, one of the problems is the bottleneck in investigator time, and so to accommodate this, we are parallelizing the entire range of applications that support the interpretation of search results.

We're a very flexible architecture and very easy to develop one. We have over 260 installations with other 100 computers in universities where research is being performed. Over 26 installations are specific to audio informatics, so there's a very good basis for this.

A result of such a large basis is continual application development and development of tools for the

processing of results.

The result of having so many bioinformatics applications running is that investigator review time can be restricted. Not only do we have the database searching tools but we have many database processing tools which is very important to removing the redundancies inherent in biological databases and also to reduce the number of results returned and save the amount of time that an investigator might have to spend going through and examining results manually.

In addition to the bioinformatics tools and computer system, Maas Parr Computer Corporation has professional services with experience in designing automated DNA processing for hybrid screening. That would mean that the reviewer time is saved both in setting up the searches but also in the interpretation of the results because much of the interpretation can be automated, and we do have experience in this field.

In addition we have educational services on the high level, even on the high level on how to use bioinformatics applications on providing search strategies and also to the lower levels of maintaining hardware, programing, language and such. Additionally, we have worldwide, 24-hour customer service support.

Now I wanted to provide some level of specificity in dealing with the problems mentioned in the announcement for this hearing, but there is limited time, so I would just like to talk about one application that's very suitable for today, and that is the clustering methodology that we have employed for biological sequences.

One of the big problems with biological databases, as mentioned before, is that there are a large number of redundancies, so therefore clustering refers to the attempt to reduce the bulkiness and the cumbersomeness of the database while at the same time preserving all the information content.

One possible model for the use of a clustering methodology is the database processing which is done at the systems support level at the U.S. PTO. This would not be done by the investigator but the infrastructures and the support mechanisms available to the investigator. A

database would be clustered, processed into another database, then having the redundancies removed and making them a smaller overall size.

The benefits are faster search time because there are less entries, but more importantly the results of such a database search would be much more easy to interpret because there are less redundancies and less multiple hits.

Another model for use in clustering is by the investigator himself or herself. Clustering can be done at any level of stringency that the investigator decides is necessary, and one possibility is doing prior art searches on multiple databases and then obtaining the results which would be very large and probably very redundant with much repetition.

The search results can be clustered to produce

processed results which will have the redundancies removed, and it will be much easier to go through. Thus, you save time in the interpretation of the results and eye strain on the investigator having to look at multiple hits of the same thing.

A very big problem in searching databases, especially EST, is that there are many ESTs for one gene, and really you only need the information of the gene itself. A clustering methodology would allow you to have one hit per one gene. That will greatly reduce the size of the databases. That will greatly reduce the amount of time that an investigator needs to interpret the results of his or her search.

The third instance where Maas Parr feels that clustering will be of use is tackling the problem of mass injuries, where instead of the single sequence, you might have hundreds or even thousands of sequences in the one application. The idea here would be for the investigator to run his or her own clustering of the data set that's provided of the patent application that's provided.

The clustering would be done at the level of similarity that they choose, and the output will be a processed submission which will contain the original sequence data, but more importantly the sequence data will

be put into work groups that are similar in pontiff levels, a similarity that's chosen by the investigator.

This saves a lot of time because instead of viewing the patent application as thousands of individual

21

sequences, he or she can view the application as a whole in its work groups and a first-level organization has been provided which will save the amount of time needed to process the application and the amount of searching that you'll have to do.

That's only one of the applications available for database processing from -- in the Maas Parr system, but we felt it was the most relevant to discuss today because it attacks the problem of mass submission. It addresses the problem of the investigator time needed to analyze the results. It even allows one to work on the database level to try to clean up databases to remove redundancies, to try to improve some of the -- some of the -- not flaws but particularities of the U.S. PTO situation.

So in conclusion, Maas Parr is striving to reduce investigator time as well as provide fast, rapid bioinformatics tools in parallel.

MR. LEHMAN: We're using your system right now. Is this the first time some of these -- that we've heard some of these suggestions, or have you been giving them to our --

MR. BURKE: This is the first time that you've heard about clustering, yes. Right now U.S. PTO is using the Maas Parr to do rapid database searching and such.

The clustering tools are rather new, but they are working currently. And they're meant for large-scale uses in database and also small-scale use an individual

22

investigator would do.

MR. LEHMAN: When you deal with the PTO, do you deal with the informations systems people primarily or --

MR. BURKE: Primarily, I believe that we deal with the information systems.

MR. LEHMAN: -- as opposed to the examining core directly?

MR. BURKE: Yes.

MR. LEHMAN: Are there any other questions?

MR. KUNIN: You began to mention a little bit about the process of post processing in addition to the two styles of clustering. The first speaker spoke in terms of setting up a process of post processing which might include the formation of an expert system which would kind of help analyze the search results from the perspective of the law, that is, to collect the information not only from the standpoint of identity but also enough structural similarity and relationships that might suggest how some of that -- some of the results might be combined for obviousness purposes. Do you have any comments from -- for Maas Parr on the post-processing solution?

MR. BURKE: The post-processing solution, we have some experience, and we have implemented similar systems in the commercial companies. It refers to automatic screening of sequences. I believe that she was speaking of a -- of a set methodology before -- sort of -- I believe she was referring to guidelines for the

investigators in doing their searches.

I think that would be very useful, and an automated schema can be put in place to save time. I personally feel that the investigator should have a latitude to perform any kind of search they want, though. I think such things should be a guideline only and not a straitjacket, but such a thing is very possible and is a very good idea, yes.

MR. LEHMAN: You know, you've heard the suggestions of the previous witness and made some of your own, and you also heard my little opening comments. And my opening comments suggest quite a large gap between the resources under our present methodology which would be required to examine these applications and what would be reasonable, given our present fee structure and resources.

Can you give me some kind of judgment on your part as to what -- how much of this gap would be reduced in your view by the proposals that you've offered?

MR. BURKE: As far as the specific amount --

MR. LEHMAN: It doesn't help us very much to -- to -- to -- to chop it away 10 percent of a problem that is -- is totally debilitating. We need to chop away 99 percent of it.

MR. BURKE: Yes. It's not just going to be buying more hardware, although that will be necessary, considering how much work you have ahead of you but also the methodologies that you'll need like this clustering.

As far as a reduction of database size, I

don't know. It won't be on the level of 50 percent. However, on the other end by the individual investigator, it could save quite a bit of time in mass applications because there's bound to be a lot of redundancy, and why do a thousand searches when you can do a hundred. I think you will need to buy more hardware. You will need to modernize, but more importantly, you'll have to start -- you'll have to keep up with the latest technology and automated processing and clustering and other such tools that can help you do more than just database searching in prior art investigations.

So you asked me for a specific number or a guess as to the amount of time that can be in saved. I don't know that I could do that, but on the top of my head, I can think about it, but I do know that it will -- you have to -- just by buying -- buying more of them will not solve the problem but employing smarter systems of the kind that I'm talking about today will -- will provide the solution.

MR. LEHMAN: Thank you very much.

MR. BURKE: Thank you.

MR. LEHMAN: Next I'd like to call Suzanne Biggs.

MS. BIGGS: Good morning, commissioner, gentlemen and lady. I'm Suzanne Biggs. I'm with the La Jolla office of the law firm of Lyon & Lyon, 4250 Executive Square, Suite 660, La Jolla, California 92037.

This testimony and comments that I'm giving

today reflect my own opinions and should not be considered as representing the position or opinion of Lyon & Lyon or of any of its clients. They reflect some of my experiences

as a practitioner in this area.

Lyon & Lyon is a law firm specializing in intellectual property law. I work primarily advising clients in patent matters, particularly in the areas of biochemistry, chemistry and biotechnology. I've been specializing in biotechnology and biochemistry for about the last six or seven years.

We understand the concerns of the Patent and Trademark Office, that they want to issue good patents that are directed to novel, unobvious, useful subject matter that will stand up to challenge in the courts. That's what we want to obtain for our clients as well.

We also understand their concern with regard to the burden of examining certain applications. However, I'm not so sure that proposing differential fees is the right way to go about it. These proposed differential fees raise three questions to my mind.

Is it appropriate to base fees on the technology type? Is it appropriate to base fees on the estimated difficulty of search by some sort of objective criteria? Is it appropriate to revise fees in a piecemeal manner when the PTO finds that a particular group of applications appear to be more difficult to examine than the norm? I believe that the answer to all three questions should be no.

With regard to the suggestion that fees be based on technology type, we believe it would be unfair to pick out biotechnology applications to have higher fees just based on the fact that they're biotechnology.

Typically, biotechnology companies are small companies. They're money sensitive. Patents are critical to their continued existence, to their continued ability to raise funds, to their ability to get their products to market. Biotechnology companies typically are one area where we seem to be extremely good at competing with the other companies in the world. In fact, so many corporate deals are being done with foreign companies, they must think that our biotech people are very innovative.

Differential fees that are much higher would

discourage patent filings. It probably would not stop them from filing the ones they see as clear winners, but some of the ones that may be enabling technology, it would encourage them keeping things such as trade secrets. It would decrease the amount of information that enriches the area in general.

Also, another thing to consider is not all biotechnology applications are large and unwieldy to examine. Very few contain thousands of sequences that have to be searched. It appears that one of the things that have really sparked these hearings are a small number of applications that seem to have caused particular problems and problems that may not be representative of all biotechnology applications.

Also, another point to consider is large and unwieldy applications, meaning those that are hard to examine, are not limited to biotechnology. Other technology areas have applications that prove difficult to examine and take a large amount of examiner time. Applications in the computer and software area and other high-tech areas and as time goes on with new technologies, I think a variety of them will have problems that way, particularly in their early stages.

As the industry progresses and matures, biotechnology applications will become more focused and, indeed, many of them are becoming more focused now. People are directing them to more specific solutions to problems. And I think this will continue to happen as the art matures.

Therefore, it is believed that it would be bad precedent to increase fees for one type of application in general such as biotechnology when there seem to be problems examining them or whenever you get a new technology type which doesn't really have established guidelines for examination.

Other fee considerations are -- many -- even focused biotechnology applications often have a large number of claims and as such they are already paying higher filing fees than many applications in the, for example, mechanical arts, where claim numbers of several hundred are

unusual.

Also, as the industry matures -- and even now while it's sort of in its adolescence and companies are looking for corporate partners to help fund clinical trials and the like -- many of these applications are losing their small-entity status, and as that increasingly happens, that will increase the fees the patent office gets, but it will also increase the fee burden on the companies.

I have a few comments to make about sequences and numbers. There's -- we've all heard some comments that possibly fees be based on the number of sequences in the application. And one thing, as the rules stand now, anytime a sequence of four or more amino acids -- as long as there are no D amino acids or 10 or more nucleotides are listed in the application, they must be put forth in the sequence listing.

In many applications, the majority of these sequences, one, are not claimed, two, need not be searched to determine patentability of what is claimed. Many of these sequences are present in the application to satisfy the requirements of Section 112: how to make, how to use, best mode. There are things like best primers. There are things like plasmid liter sequences, things that the applicant does not expect to get patent coverage on. Often, they will be things that if anyone would be entitled to patent coverage, it would be a company that, say, provides primers, plasmids and the like. So we believe that it would not be appropriate to base these fees on the number of sequences.

We can charge higher fees just because there are a large number of sequences in the application, where these sequences need not be searched in order to determine whether a patent should issue on the claim subject matter.

In fact, complying with the sequence listing requirements can be fairly burdensome in some of these applications just because of the number. We can understand the importance because you all, however you decide to determine your database, need to have all these sequences that are known in it where appropriate, but for determining

whether a particular applicant is entitled to a patent, they're just plain not germane.

I've heard several suggestions for easing the alleged problem of searching sequences. At present, it seems to be limited to a relatively small number as compared to all biotechnology-type applications but very large applications. And perhaps guidelines can be made to determine what a common focus should be when multiple, large numbers of sequences are being claimed, but that's something for those who are more involved with those sorts of applications to deal with.

There's also been some suggestion of applicants making a prefiling search and putting forth some sort of certificate. There is a provision now in petitions to make special, where applicants are requested to put forth -- one of the grounds is a prefiling search and various representations.

In view of the sorts of representations and the uncertainties of how those would be construed in a

later proceeding, I see in my own experience very little use of those provisions. And with respect to a prefiling certificate of some sort of search by applicants, I see some problems.

Those problems would be: and how do you standardize what has been searched? What databases do you have to say this search was done within a certain number of days before filing? Also, in determination of guidelines, how close does something have to be before you have to note it in your certification?

The other issue that I can see is a problem is: Who would make the certification data that a particular search that fits certain criteria is made? Would it be the attorney? If it's the attorney, that would further increase costs for the small clients because it would require a fair amount of attorney time to review the search and the protocol.

If it's a nonattorney, how would we know what the quality was of the people doing such a search, that they really understand the databases, they understood the criteria? I can't -- these are just issues to consider in

going forward with any sort of the suggested changes, and I thank you for your time and your consideration.

MR. LEHMAN: Thank you.

Are there any questions?

MS. LINCK: I have one question.

You mentioned that we may not search all the sequences disclosed in the application, but what happens

then when the claims are amended to claim one of the sequences that wasn't searched initially but was disclosed?

MS. BIGGS: You would have searched at that time the sorts of sequences. I was particularly commenting on, for example, you have an application where you've isolated a new protein from a source. You have cloned the protein. Many of the sequences that are involved in the cloning are probes, primers, things like that that under these regulations in the 37 CFR 2nd, 821 and so forth will have to be listed. However, these are things often that are either known sequence or they're things that you just are not claiming. They are just not related to what is being claimed. They're not related to the proteins, sequence of the protein that are being claimed. They're not related to the CDNA.

MR. LEHMAN: But we would have to identify -- these are in the prior art already, which is what the search accomplishes. Would you offer an admission in the application that these are prior art sequences?

MS. BIGGS: Well, if we're not claiming the sequence, many times we will say where the sequence comes from, that they come from a particular provider of liters that one puts in a mass mix or they're just things that we don't claim. Of course, if we later claim them, they would have been to searched, but just as one pays if you start out with 20 claims and you add another 40, you pay for them. If one were to amend one claim to add an additional 50 sequences that the Patent and Trademark Office could do

if they were to base a fee on the number of sequences that were claimed as opposed to presently the application they'd deal with that, but many of those sequences don't need to be searched.

MR. LEHMAN: So would you support an increase in fees if in fact an amendment would be made that required additional searches to cover the additional costs of the search?

MS. BIGGS: Right. Now, first of all, I don't support increased fees for biotechnology applications. I would have to see some sort of proposed regulation. I think in general where they're focused, it's without a full review of all sorts of applications that may be difficult to search. I don't think it's fair to pick out one area of technology as having -- that may be easier to quantify. It would be like in pharmaceuticals if you have a group that claims more than 1,000 compounds, you have to pay extra. And some of those, I'm sure, must be very difficult to search appropriately because of -- even with the restriction practice as it is now.

MS. LINCK: But if there was a method for adjusting fees to make it fair across the board, have the fees more aligned with the actual costs to the office, would that be something you would consider?

MS. BIGGS: I'd have to see a concrete proposal I think.

MS. LINCK: Thank you.

MS. BIGGS: Thank you.

MR. LEHMAN: Thank you very much.

I'd next like to ask Richard Klobuchar, please.

MR. KLOBUCHAR: Good morning, commissioner, members of the trade office and members of the public. My name is Dr. Richard Klobuchar. I'm currently vice president of the corporation of Science Applications International Corporation. SAIC is headquartered here in San Diego. My office is at 200 North Glebe Road in Arlington. I'm currently also the chief scientist for SAIC's Systems Engineering and Technical Assistance support contract for the Patent and Trademark Office and currently co-principal investigator for a pending distributed supercomputer project with ARPA and the PTO which is referred as to the Metacomputage Project. These comments represent the comments of my corporation.

SAIC is an employee-owned, diversified, high-technology research and development services company focusing on information systems, telecommunications, security, transportation and health. Our 22,000 employees are organized across autonomous sectors and transform emerging information technologies into customer-focused solutions for literally thousands of government and commercial customers. Under the competitively awarded System Engineering and Technical Assistance contract for the United States Patent and Trademark Office (USPTO), we have been performing such technology infusion tasks since June of 1993.

Since September 21st of 1995, SAIC, as the PTO-SETA contractor, has been supporting the USPTO in the selection of hardware, software and approaches to facilitate their examination of biotechnology patent applications. Efforts to date have focused on the analysis of requirements and has resulted in a Preliminary Molecular Sequence Searching Systems Requirements Analysis document dated December 19th, 1995.

We anticipate continued efforts for the USPTO through a new Task Order entitled "Automated Support for Biotechnology Patent Examination." Under this Task Order, the SAIC PTO-SETA staff will provide technical and analytical support to enhance USPTO capabilities for automated biosequence searching in support of PTO examiners.

On the 15th of March 1996, SAIC's Laboratory Sensors and Automation Division, located here in San Diego, entered into a contractual relationship with Incyte Pharmaceuticals, Inc., for the development of a next-generation robotic system in support of Incyte's DNA sequencing operations.

On the 18th of March, SAIC notified the USPTO that via the Internet, we had discovered the work being performed by SAIC's Laboratory Sensors and Automation Division in San Diego. In our OCI notification letter, we described the work being performed by SAIC's Laboratory Sensors and Automation Division. We went on to explain

that the role of that laboratory is strictly in the area of robotic enhancement for the Incyte laboratory equipment and, therefore, we felt that no Organizational Conflict of Interest (OCI) existed. Nonetheless, concerns have been raised by the USPTO that there is the potential for a perceived OCI by the public, especially for those corporations who compete with Incyte in the filing of biotechnology patent applications.

As a result of these concerns, we agreed to two actions; one, to develop an OCI Mitigation Plan and, two, to make a public statement, this statement, regarding our activities. The OCI Mitigation Plan was delivered to the USPTO on the 2nd of April. That plan fully identifies and addresses all possibilities for a perceived OCI in task areas pertaining to support for biotechnology patent examination. Our plan establishes Organizational Controls in order to isolate the SETA contract from all the other contracts being performed by SAIC, both physically and operationally. Further, the plan establishes an Information Control System consisting of Physical, Data, and Administrative controls in order to protect sensitive information from public disclosure. And, all SETA personnel assigned to any biotechnology-related task will receive special, targeted OCI training in order to reinforce the execution of this Mitigation Plan. We believe that this Mitigation Plan addresses USPTO OCI concerns, ensures the confidentiality of USPTO information, and has removed any remaining OCI obstacles for the continuing PTO-SETA support. Thank you.

MR. LEHMAN: Thank you.

Are there any questions?

Ms. Kepplinger, do you have the other witnesses who signed up?

Next I'd like to call Barbara Luther, please.

MS. LUTHER: Good morning, gentlemen and, Mrs. Linck. I am Barbara Luther. I am here to represent the biotechnology industrial organization. My address is at Incyte Pharmaceuticals, 3174 Porter Drive, Palo Alto. These comments are the summary comments from Bio patent committee meetings. There were a number of other

suggestions that I won't mention today that we could not come to agreement on, but we certainly appreciate the opportunity to come and talk to you again.

MR. LEHMAN: This reflects the Bio Industry Trade Associates Patent Committee?

MS. LUTHER: Right.

First of all, I think we have to look back to October 1994 and applaud you on the progress made. We are also here to propose future goals and even some specific suggestions on how to accomplish them.

Biotechnology comes of age: 11 major biotechnology products qualified for marketing in the last year alone. The list of problems solved by biotechnology drugs grows longer.

First of all, recombinant insulin, a protein manufactured for diabetics; recombinant human growth hormones to help children grow to normal size, recombinant

hepatitis B to protect people from liver destruction; recombinant t-PA to stop heart attacks in their tracks and return people to productive lives in days, as opposed to weeks; recombinant Factor VIII to stop the crippling bleeding in hemophiliacs -- a pure drug which alleviates obtaining Factor VIII from the commercial blood supply, which formerly resulted in many bleeders getting hepatitis, then AIDS; Epivir for AIDS; Doxil for Kaposi's sarcoma; Vitasert for CMV retinitis.

To fund this tremendous research effort, companies have needed the promise of getting a patent to protect their efforts and give them enough money to pay back their investors. The companies themselves only seek to patent isolated and purified DNA and proteins, as they exist in our laboratories after the hard work of our scientific research teams. Patents did not cover DNA in your own body. Patents are only wielded against competitors which try to profit from our inventions.

Many more biotech products are in the works and in the news. The early drug pipeline has literally of hundreds of drugs in Phase I and Phase II studies.

Many new products and services are used in the

emerging biotech market arena. It is highly computerized. As you heard from our gentleman at Maas Parr, now sequencing machines can prepare sequences for easy loading into computers. Computers programs have been designed to compare sequences and permit the scientist sitting at the computer terminal to make realistic predictions about the

38

structure and function of recombinant protein. Computer-modeled drugs are going into clinical testing.

Companies interested in leveraging biotechnology expertise can turn to a number of genomics companies, including my employer Incyte Pharmaceuticals, to out-source their biotechnology discover. We genomics companies provide new therapeutic proteins, diagnostic probes, human receptor targets for drugs. Breast cancer and obesity genes are examples of Incyte's competitors' successful collaborations. But enough of our success, let's look at the PTO accomplishments in detail.

In October of 1994, you took quite a beating over the utility issue. At that time, you insisted that you wanted to give us quality patents we could take to the bank. And you instituted changes. You issued guidelines and legal rationale for changes. The PTO followed up to assure that these changes would be implemented by the examiners. You drafted explicit examples which showed examiners how to reject claims but also how to allow claims when we met the appropriate standard of proof. You implemented quality control and achieved value on the utility front.

Another great accomplishment that is just beginning, the patent office has been training a new crop of examiners to propose allowable claim language on the first action. This revolutionary change, if sustained by the office, could save biotechnology companies millions in patent attorney fees, and these millions will be used to

39

further research. This would eliminate unnecessary narrowing of claims. It would eliminate prolonged file history battles. It would eliminate some of the problems we have encountered in celebrated cases such as Genentech versus Novo Nordisk.

A third accomplishment, you also appointed

biotech practice specialists, who monitored the performance of individual examiners and specifically retrained examiners. Though I haven't had to resort to them yet, they being there holds great promise in helping assure quality and fairness throughout the biotech examining group.

And two weeks ago you demonstrated, for the first time in my memory, your sequence searching techniques. You have developed a fine system as the art and science have evolved. You are being complimented on bringing your sequence source problem to the attention of the public so we as customers can help.

Future issues. Sequence searching. This is a really big problem. We hope to continue to work with you to build a fair and efficient searching technique. These are difficult and complicated computer techniques and bioinformatic techniques. They can help make your job a lot more efficient. We would like to see you craft an internal quality standard or set of standards for searching, much as you did with utility. I will provide some detailed suggestions on that in a moment.

Another big issue that you need to address is

the scope of claims for patents with DNA and amino acid sequences. It varies considerably. Some examiners present precise sequences. Others will allow broader scope. In fact, it is relatively easy to substitute DNA codons, as 1 to 6 codons specify each amino acid. Similarly, conservative substitutions of amino acids have been known for more than a decade; one amino acid with another comparable one, so that there is no change in structure or function. This state of recombinant technology art is such we must change the human sequence for more efficient cloning in bacteria.

There are computer programs now which will convert an efficient human gene into an efficient E. coli gene. If the examiner limits the claim to a particular DNA sequence, competitors could avoid such a narrow claim. If we are to have a patent we can go to the bank on, we need adequate claim scope. We propose continuing discussions on

this topic.

Section 112 issues: Sometimes it seems like examiners have converted their utility rejects from "It's incredible that you could do this" to "You haven't taught me how to do it." If we cannot get a patent, we'll still be delayed years in getting biotech patents. The development of drugs today is literally an industry, with regulations that stipulate when and how to do certain testing. Pharmacologists are familiar with establishing dosing, whether for antibiotics or other recombinant proteins. Obviously, the examiners had experienced drug

developers, and so those topics are largely unknown to them, but not to those familiar with the current state of the art. Again, this is something we need to work on together. Perhaps I should go back to B.I.O. and see if members would have scientists to provide seminars on these topics.

Now, for some specific steps. Step one, I don't know if you realize this, but the examiners are asking the inventors to send in any sequence that appears in the patent application literally. If we have a new hoopla of human growth hormone factor and we show a comparison in the figure, they tell us to send in a sequence listing, the sequence for the known human growth hormone factor that already appears in Genbank and is no doubt in your database already. And we see these day to day. We know that your database may be very redundant. We could work together on this. If we simply cite to a known database location, called G.I. number for Genbank, that should be sufficient.

Step No. 2, because of this redundancy, we know that you already have a redundant sequence. So as the gentleman from Maas Parr said, we need to work on that redundancy. It may be double, triple, quadruple the sequences it may be already. And from the pace the research is going and the pace it looks like, you'll be receiving a lot more sequences in the future. Streamlining your database will be essential.

No. 3, convene a meeting on the sequence

analysis technology with industry, academic and government

experts. Definitely Maas Parr has some new information. And hopefully working together, we can get to the goal of helping you search. It would be helpful if we started an information exchange, which sounds like it could be beneficial, and I know you'd like to continue that on an open forum, too.

The state of the art does appear to offer software tools and help subjects to come up with all the relative but not the superfluous prior art. From the PTO presentation, it appears likely that inefficient searchings is excessive and bogs down examiners with excessive printouts and diskettes.

Although such a meeting should take place as soon as possible, there's a big bioinformatic meeting in Baltimore in early June. Maybe your schedule would allow so we could help arrange or follow on or present a meeting of that. All the experts are meeting in the neighborhood at that time.

Step No. 4. Draft internal quality standards for sequence searches by patent examiners just like you did with the utility examiners. We should have an appropriate search standard, then we'll get something we can go to the bank on. Providing these examples to the public will encourage biotechnology inventors to ask for rational coverage and support it with the searches they routinely perform.

And Step No. 5, after we get that expert input

and quality standards, let's reevaluate the cost to do searches and examiners' time. And then that will be a good time to have a meeting and again request the investors' input to help. That will be the time to discuss additional approaches, such as providing tailored searches.

Again, we applaud the progress that you've made. We really appreciate the way you've been working with us, and we look forward to cooperating on these initiatives in the future.

MR. LEHMAN: Thank you very much.

Any questions?

I have a question relating to Incyte. I

appreciate what you had said about the utility guidelines, and obviously as ICF being the essence of utility guidelines, that you have to show that there is going to be some utility, for example, in the case of biotechnology inventions, that relates to the pharmaceutical, ultimately what would be the disease that would be cured and so on and so forth, and you have provided four mechanisms far short of human clinical trials to show that utility, and so on and so forth, but you've indicated Incyte is a -- I think you referred to it as a genomic -- what was the word?

MS. LUTHER: Genomics company.

MR. LEHMAN: Genomics company. And the applications that we're talking about are largely applications that are flowing out of genomics companies rather than other kinds of biotechnology companies. And so first I'd like you to explain for me a little bit more what

the difference is between a genomics company and other people in the industry. Can you do that?

MS. LUTHER: We have Randy Scott here who is our vice president and chief technology officer who does an excellent description of that.

MR. LEHMAN: He's on the list?

MS. LUTHER: Yes.

MR. LEHMAN: I'll ask him then, since he is going to be preparing an answer for that question. And maybe I should just wait until he comes to testify to follow up on that line of questioning.

MS. LUTHER: We believe by issues of state of the art bioinformatic and using the biology, we do know about the sources of its sequences and the multitude of comparison that we can make among all different libraries that we have representing multitudes of different tissues in the human body, that we can come up with a good argument for utility for many of the sequences.

MR. LEHMAN: Thank you. Are there any other questions? If not, we'll move on to Randy Scott.

MR. SCOTT: Good morning, my name is Randy Scott. Hopefully a good part of my comments will address exactly some of your questions and issues. I would really like to structure this in two different parts. One, first

of all, is just to step back and talk a little bit about what we think is the future of the vision of this whole field of bioinformatics and genomics because I think if you want to deal with practical problems today, it's incredibly

45

important to be envisioning what this world is going to look like in two years so that you don't have to be solving today's problems two years from now and realize that all you've done is put yourself behind the eight ball in terms of a whole new set of problems. So I'd really like to talk a little bit about the vision that Incyte has and the way that we believe the world is going to look in a few years because of bioinformatics, genomics, which is now a very, very important and vital sector of the biotechnology marketplace.

To sort of give you an introduction, I'd like to go back about a year ago when I was with my son, who is an 11-year-old, and we went to the movie "Apollo 13," and we were sitting there watching the point in "Apollo 13" in which the projectory is off coming back into Earth, and so they had to change that trajectory slightly. And we all sort of think of NASA as this fantastic, high-tech organization, and I was amazed in the movie, which was quite accurate, in which the -- one of the astronauts pulled out a slide rule and calculated the trajectory on a slide rule.

Of course my 11-year-old son punched me and said, "Dad, what's a slide rule?" because he's been playing on computers now for most of his life and using calculators, and so he had no recollection of the level of complexity that he has to work with in today's environment versus what they had in the 1960s. And it struck me right as he asked that when I was growing up in the '60s, my

46

parents and everybody at that time used this phrase, "Why is it that we can put a man on the moon, but we can't cure the common cold?"

And it sort of struck me right at that moment that the reason is very simple. The complexity of the human body and of biology and disease is in fact millions of times greater than the complexity of putting a man on

the moon. So I think what we've looked at in medicine as a very high-tech field and sector for a lot of years and we've patted ourselves on the back, in fact, we at Incyte believe that we're in the very, very early stages -- in fact we're not high tech at all.

Imagine trying to play a game of cards with a deck of cards in which you only have five of the cards to play with. There's not many games that you can get involved with. And in fact what Incyte is doing and what the worldwide genome effort is doing is first and foremost just trying to take apart the human body and figure out what are the 100,000 genes that make us us. And in fact the mechanic of your car is not going to do very well if he doesn't even know all the parts that exist in your car. And so until we have all of those parts and all of those pieces in place, our fundamental understanding of biology is always going to come far short of the absolute complexity with which it exists.

Just to give some examples, most of the drugs that are on the market today were probably developed 18 or 19 years ago at which time maybe only a few hundred genes

were even identified and characterized to any great extent.

As of 1990, we estimate there are probably only 2,000 genes within the genome for which we had a complete structure, much less much in the way of downstream biological information. And yet in the last five to six years, because of the advent of hybrid dna sequencing technology, we've now been able to identify at Incyte what's probably approaching 100,000 genes within the human genome. In fact our expectations are that the level of complexity we're still seeing, that there may be as many as 150,000 genes or more within the human genome.

When you think for a second that means in five to six years time, we're going to go from 2,000 to 100,000. That's almost two orders of magnitude increase in the number of genes we now have available for researchers to study. And so I think throughout the world, that's been brought up as an enormously complex issue. We have to be able to deal with these types of problems, not with paper lists but rather with very sophisticated computer analyses

to be able to help us to examine that data.

But in fact that's only the beginning of the complexity. Because the second order of complexity is not just having a list of all 100,000 genes in the genome, which in fact we will have and many different companies and the public domain will have within the next few years, but knowing which of those genes is expressed and turned on and off in different phases of biology.

Now, if you take my prediction of 150,000

genes within the genome and you just take the position that each gene is either on or off in any given cell or tissue or biological environment, that already gives you any number of possible expression states of the genome of 2 to the 150,000th power. While that number is not infinite, but it might as well be from our practical aspects of how we can look at that data and yet that's precisely the goal of companies like Incyte Pharmaceuticals, is to understand not just the list of the 800,000 genes in the genome or more but also which of those genes -- and typically it maybe anywhere from 10,000 to as many as 30,000 genes that are expressed in a single biological setting.

We don't want to know just those genes that are expressed in the central nervous system. We want to know the genes that are expressed at a particular time and state of the central nervous system that might be associated with disease. So as you begin to think about the complexity that I've described here, you begin to realize that computer tools and informatics are absolutely the only way that we'll be able to understand this complexity, so they're absolutely pivotal to how we move forward as an industry and I think how the patent office is going to be able to deal with the increasing complexity of pharmaceutical inventions because we're going to be providing you with data not from four or five genes, which we've measured by a lyse analysis that predicts that one of those genes is associated with, for example, Alzheimer's disease or rheumatoid arthritis, we're going to be

providing you with electronic, computerized information, databases that suggest we've done exhaustive studies of

these 10,000 genes, and we can show you which of these 10,000 genes are up-regulated during this biological event and which are down-regulated. And we can correlate that with structure and with function. And we can make very good predictions about which of those molecules are going to be important, not just to cure the disease but to understand the disease to begin with, which we think is one of the primary reasons for our existence, is to be able to identify markers, genes that could be used either in a clinical setting, in a diagnostic setting to look at a particular disease or as clues for the pharmaceutical industry to be able now to exploit some cured disease.

Genomics, in fact, we believe -- or I believe -- is simply a commercial term that is risen to describe what the public sectors called the worldwide human genome effort. There are two different parts to what I believe are genomics. The first part is hytrobid [sic]tools for studying biology thousands of genes at a time. And in the early days most people recognized that as hytrobid [sic] DNA sequencing. We have come up with some smart ways to do the DNA sequencing, some very sufficient ways to do the DNA sequencing that allow us to identify expressed genes in a very, very rapid rate. But that's just the beginning.

Once we have each gene, when we sequence it the second time, we can identify the gene that we've already seen, and we can be able to compare where we've seen that gene before. And so we're not just doing this at the genomic level. We're doing it at the level of biology. We're understanding which gene is terminal and in a prostate cancer sample versus a normal part of that same prostate just a few millimeters away, being able to do an exhaustive comparison of those two samples of which genes are up-regulated and down-regulated which leads us to a fundamental understanding of which genes are involved in the prostate cancer process. So that's the first step deciphering DNA sequencing.

As you may know, there are many, many technologies that are being developed now which will allow us to take any excessive genes that we can identify on a

database, put those onto literally a silicon chip or a glass slide and to be able to rate thousands of clones and now to be able to process multiple biological samples across those thousands of clones.

So now after we do our initial studies, we can take a thousand genes that we believe are involved in prostate cancer and we can take biological specimens from a variety of different individuals, and we can test that hypothesis, and we can demonstrate further with clinical samples which of those genes are regulated only in a portion of prostate cancers and which are regulated in all prostate cancers. And in fact in many diseases, we'll find that there may be no single gene that correlates with that disease but rather patterns of different molecules of

different family's members which correlate with that particular disease, and that will allow us to explore that information further. In fact we believe that that's the future of diagnostics, is not one molecule at a time but thousands of molecules at a time.

How would you look at that data? How do you look at data derived from searching 200 patient samples across a thousand or more molecules?

A simple list or even a computer printout profile is not enough. You have to be able to use very large, very sophisticated relational databases to store and allow you to search and retrieve just the bits of information that you want to utilize for the particular studies or for the particular analysis that you want to do, which brings me to what we believe is the second part of the genomics effort is not just creating a hytrobid tools for analyzing biology thousands of genes at a time but information tools and databases, tools for mining those databases that allow us to get the unique data sets that we need very, very quickly.

This is not unique to biology. It's not unique to bioinformatics. This is a wave that is taking over the entire economic sphere of our country -- database mining. As we're able to store larger and larger amounts of information, the tools now for mining those data sets

and being able to create usable information are enormously valuable. And I would submit that one of the things that the patent office needs to be thinking very seriously about

is moving towards electronic filing.

As we file large number numbers of patents with large numbers of CDC sequences in those patents, that paper stack cannot even begin to explain the complexity of the database from which it was derived and our ability to search that database to compare the sequences against each other and get results within seconds, not hours or days. And our ability to compare where each gene and each individual biological setting is as expressed, that information is only going to explode over time.

Our database currently houses several megabites of information, but we see that whole field expanding by orders of magnitude in complexity over the coming years. You never use all of the data all the time. No biologist is trained to be able to sit down and use all of that information. You're only able to mine that information for the observations that you want to follow up on.

So the implication for biologists, and I think for many scientists, is that our whole world is going to be changing over the next several years. We don't think bioinformatics is simply a tool that a few specialists will be using. We believe it will be the primary basis in which biologists will be able to retrieve, do searches, analyze and set up their individual experiments.

At Incyte, we draw a lot of analogies for our business to the computer industry with its emphasis in both hardware and software. For us, hardware means that first

part of genomics, hytrobid tools for analysis. Our hardware is DNA sequencing and tools for analyzing DNA samples or sequences at a very, very rapid rate.

Today at Incyte, we're sequencing between 6,000 to 7,000 cDNA clones per day. That's an enormous array of sequencing. In fact, we believe that there's probably a corollary between Moore's law and the semiconductor industry and what we're doing today in DNA sequencing. That is, we double the rate of sequencing at

Incyte and have doubled it almost every six months since we started the company down this path in 1991.

If I'm correct with that, that means from 6- to 7,000 sequences that are being analyzed a day, we'll have to process 12,000 or more in six months, 24,000 per day six months after that. And algorithmic expansion, of course, means we have to build all the tools to do that. In fact I would argue that we've done that in a very, very satisfactory fashion. So let me just give a brief explanation of the bioinformatics tools that we use in house.

And I think the philosophy or the theory comes back to database mining: Don't try to create a large set of data about every single molecule that you're going to look at. Select your questions very carefully. We tell our scientists all the time, "Be careful what you ask for because you may get it." So if you ask for an all-encompassing report of every sequence, you will get that, and you will be so overloaded with information, you

could never analyze it, and the whole system will grind to a halt.

So in our system we take some very straightforward approaches. First, do the easy searches first with the fastest tools and the easiest tools, ones that don't take the computer intensity of time. So within our process we can search and analyze about 6,000 sequences a day in a pretty straightforward process in which, first, we take every one of those sequences, and we go through a series of masking programs, and we eliminate all the things that we know we're not interested in like repetitive DNA elements, alpha sequences, things that are not useful information because they've been sequenced many, many times over -- there's nothing inventive about that process. But there's a key reason to do that is because when we now turn around and search our sequences against the public domain, if we don't remove the so-called junk sequences out of our information, what we'll find is that our junk will match up with public domain junk and you end up with junk squared. And that's not good. So you want to get rid of at least --

have at least one data step in your comparison set in which you've removed all that and we go so far as to mask things like dinucleotide or trinucleotide repeats, so we get down to the sequences, just the things that we're really interested in.

Then we go --

MR. LEHMAN: How do you distinguish between the public domain and what is yours?

55

MR. SCOTT: Well, that actually comes in a second phase of the analysis. In the first phase of the analysis, we've worked with numerous academic collaborators, probably many of the same people that you would find available for consulting roles to set up databases of all of the mitochondrial genome, the ribosomial genome, almas, repetitive almas. We've established what those outside collaborators -- Temple, Smith and Boston University was one of the founders of Genbank, Jean Claude Raboret [sic] of the CNRS and many, many other scientists' parameters that are fairly widely agreed upon in the bioinformatics community to be able to eliminate those elements first.

MR. LEHMAN: Well, is your distinguished -- is what you distinguished between being the public domain and what is yours then simply what has previously existed --

MR. SCOTT: That's --

MR. LEHMAN: -- prior to your --

MR. SCOTT: That's correct. That comes in the second stage -- phase of the analysis. Once we've eliminated those, probably for our purposes, more uninteresting molecules, we do a very fast search, which is called a blast search. That's -- we can set up multiple sequences processed in batch mode. It's not the best searching algorithm, but it is the fastest. And you can search huge volumes of sequence information very, very quickly. And the only question we're asking is: "Do we have an exact match to Genbank primate?"

56

We don't need to look at C elegance. We don't need to look at yeast. We don't need to look at bacteria because we know that our feed sequence comes from humans, and so we simply look first for an exact match in a very

small portion from Genbank. And over half of our sequences are immediately caught through that process as exact matches. And so we can eliminate those right off the bat.

And then we can take the remaining sequences and just start to do more in-depth searches. We can say, "Okay. Now let's look for homology in the human primate database or" -- first what we do is we go to rodent database and we say, "We have something that is very closely aligned to a mouse or rat sequence" because, again, that allows us to identify "This looks like the human correlate to a known mouse or rat gene."

Then we can go back and use more sophisticated tools to look for homologies. And I would suggest that that in fact is one of the approaches that the patent office would want to take, is to do the easy things first. Come up with ways that will allow you to screen and sort of immediately eliminate half of your work load or more. Then only come back to the narrow subset of things for which you want to go in much greater depth and use the more sophisticated algorithms such as the Smith-Waterman algorithm I believe is running under Maas Parr instruments to do that second process.

So by going through these stages, Incyte actually does, I think, a very high-quality job at cleaning

up, comparing things to the public domain -- first simply identifying what's an exact match and what is not and then going through a process of trying to annotate those molecules that are not exact matches to show what are they most structurally related to and then from then on we can select from molecules that we think are highly related or are highly up-regulated in certain biological cases to really hone in on the individual value of those molecules.

That system is now in place and has been running since September of 1995. So well over six months it's actually been put in place in collabortion with the collaborators, as I mentioned, many of them who are free to consult and work with other groups and in fact we would quite encourage you to have some of those groups in because our daily life is surrounded by, you know, the process of

information. We don't do it -- if we don't do it in a rapid time frame, we won't exist because we can't present this information in our database for pharmaceutical partners. So I think I can probably going on for some --

MR. LEHMAN: I'd like to ask. What is -- how long has Incyte been in existence?

MR. SCOTT: Since April 1991. How far along was our corporation.

MR. LEHMAN: And do you have sales right now?

MR. SCOTT: We do have sales. We have six database subscribers, which are some of the six -- some of the best pharmaceutical companies in the world -- Pfizer, the pharmacy at Upjohn, Nomen Ordiz, Johnson & Johnson and

Abbott Pharmaceuticals. Our current annual revenue rate from sales of the database is in the range of $25 million per year. So in fact as I look more and more at the software and informatics sector, there's an awful lot of software and database companies out there from other fields that don't yet add up to 20- to $30 million in revenue.

Our marketplace, we believe, is the top 50 pharmaceutical companies in the world, to get this information into their hands as rapidly as possible. They have accessed our intellectual property portfolio. All of those companies have nonexclusive access to the use of our sequences for research purposes and then exclusive access on a first-come, first-serve basis to individual gene product patents that might derive from our work. So in fact we are creating sort of a unique resource for the pharmaceutical world. And in many ways, patents are almost more important to us because what we're selling is ideas.

MR. LEHMAN: When you talked about how you do your work, you said that you start out comparing your genomic mapping to what's in the public domain already, and you identified that as basically work that's already been done by other people, largely academic scientists.

What kind of reciprocal relationship is there between you and them? Let's say, you know, the scientist at Harvard wants to follow up on research that he has done. Do you have relationships with them? Do you share information with them? And if you do, what kind of terms

and conditions do you apply to them?

MR. SCOTT: We actually do so, but we do so indirectly because we don't really regard this database as ours. We regard it as our pharmaceutical partners. So almost all of our pharmaceutical partners in fact have programs by which they interact with the academic community. We quite encourage that. So while we do have some individual academic collaborations -- one, for example, is an institution by the name of Mayo Clinic by which their researchers at the Mayo Clinic can get access to full-length clones that derive from Incyte database through collaboration there. Most of our academic collaborators go to the pharmaceutical companies themselves because it often turns out that they, like we, are looking for capital to fund their research. And capital, you'll find, is not within the biotechnology companies but within the large pharmaceutical companies that are really developing and dropping off of this business.

So we think in fact there is a tremendous synergy here. In essence, the pharmaceutical industry is paying Incyte to very rapidly try to create some of this information ahead of time to come into the public domain. And so by funding that program, they're both accelerating the pace of research, and they, as they look through the database and work with their own academic collaborators, can identify individual molecules that they now want to pursue more aggressively and encourage --

MR. LEHMAN: Would you say that what -- primarily what you're doing is creating a database on a

human genome?

MR. SCOTT: Absolutely. I think Incyte is an informational technology company.

MR. LEHMAN: And that database is used as a tool, or when you speak of "tools," are you speaking of something different from the use of the database?

MR. SCOTT: No -- well, I think the tools are many fold. They're not just the database. They're -- they're the database, the proprietary software we've created around the database to be able to help them mine

hat database and the actual molecules themselves. Because what's unique about Incyte is that not only can you search through the database to find a molecule, it's not coming out of an archive like Genbank, where now you have to go track the original reference source to find out if they're even willing to give you material. Our pharmaceutical partners immediately have access to every gene for which we have a cDNA stored away that we provide within days after that. So they're also requiring access to new receptors, new targets, and they have access also to the intellectual property portfolio and the molecules for which we're filing patents on.

MR. LEHMAN: Are you filing patents in other countries besides the United States?

MR. SCOTT: In other countries, we're filing patents on full-length clones only and being very selective about how we filed on EST sequences, primarily because of the publication rules because we believe that our economic

value and what we're being paid for primarily by the pharmaceutical industries is for lead time. And that's true of almost all pharmaceutical research. They know that ultimately every pharmaceutical company will have the same genes and the same products, but if they could get a lead time past their competitors, then they can move very, very quickly. So anything that of course puts all of our information into the public domain simply means that the capital resources for doing this will dry up because --

MR. LEHMAN: One of the -- you -- sounds to me like fairly accurately compared what you're doing to the more classical software industry and that in fact there is sort of a merger going on here between software and biotech industries and what you're doing, and of course the software industry has for a long time kept the light on multiple forms of intellectual property protection and not simply patent protection, to rely on copyright protection as well.

Copyright does not protect what is just pure information, however, and thus the Supreme Court has decided that pure information in the form of database -- a database -- even with that database is very difficult to

create and requires a lot of investment that has no protection.

This is a global problem, not just a U.S. problem. And in fact the European commission is in the process of promulgating directive to all the members states to the European union to enact laws which would provide a

62

new kind of intellectual property protection for -- specifically for databases, even though they don't meet the test of what we call originality in our copyright sense.

MR. SCOTT: Right.

MR. LEHMAN: But it would be more of a copyright-like type of protection as opposed to a patent type of protection. Really protects the information in the database and from unfair extraction or unfair exploitation of that information.

Would that be something that would make your life -- and it's easier -- it would be easier to get, by the way, because then you wouldn't have to apply for patents and, you know, spend all the money and spend a couple years trying to deal with the issues that we're talking about here today. Would that sort of protection be something that would be welcomed for you?

MR. SCOTT: I think on the databases as a whole, certainly would allow me to sleep easier at night to know that we have some level of protection should that database ever be stolen or, you know, other sorts.

MR. LEHMAN: Particularly important in a computerized environment because in the old days when you had databases that were contained on paper basically, you know, just physically there was a certain protection. You know, you couldn't --

MR. SCOTT: Right.

MR. LEHMAN: -- you couldn't go in and copy stuff out of it but obviously now where databases are

63

available electronically, you can -- if you can enter into the database and enter into the security that may have been provided by incryption, you can, you know, extract information from it, and of course nobody knows where you got the information.

MR. SCOTT: I think absolutely that would be welcomed. It probably wouldn't change the patent stance because, I mean, we really are a merger between computer science and microbiology, and the value, the information now leads you to which molecules you want to take forward into product relationships, whether it be diagnostics or therapeutics. So I don't think it would be looked at as a solution to the patent dilemma of the patenting of thousands of molecules at a time. And the other thing, this dilemma that we're now seeing in the patent office with the number of molecules that are being discovered is really -- it's the embarrassment of riches or the fact that such a wealth of information can be created, everyone knows it's so enormously valuable in such a short time, I think, is much of the -- much of the shock.

So still individual molecules are something that we would continue to pursue. We have a whole group of biologists that we call the Group in Silico, biologists -- for years people have done in vitro and in vivo biology, we think the state of the art now is so good with bioinformatics that -- certainly not for all, not even for most, but for a large number of molecules there is a clear structure-function relationship that's been hammered out in

the laboratory. And now we can identify members of this family. And even more importantly than that, it's highly up-regulated in the joints of arthritis patients. So we believe this is a clear-cut observation and be able to provide this molecule in a way in which the pharmaceutical sector can both have intellectual property protection on the molecule and be able to pursue those aggressively, I think, is going to be tremendous value to society.

I think people will be shocked at the amount of knowledge of disease processes, of the most complex disease processes we haven't been able to get at by doing biology one at a time like cancer, AIDS, writing different viral settings as well as arthritis. These are complex diseases probably with multiple etiologies, multiple genetic makeup and background and will not fall into one disease category. And it won't be until you break those apart into 10, 15 separate disease categories and different

treatments and regiments that we're really going to be successful in curing them.

So I see tremendous value added over the coming years, not just by Incyte or our colleagues and competitors in the genomic sector but now as pharmaceutical companies really come into this field in a big way to be able to exploit the information and develop drugs.

MR. LEHMAN: Are there any other questions?

Thank you very much.

Next Michelle Duran.

MS. DURAN: I signed by mistake.

MR. LEHMAN: Oh, so you're not planning to appear.

Ned Israelsen.

MR. ISRAELSEN: Good morning. Thanks for the opportunity to testify at these hearings. My name is Ned Israelsen. I'm a partner with the law firm of Knobbe, Martens, Olson & Bear. I head up our biotechnology group. The address of our local office here in San Diego where I work is 501 West Broadway.

And I'd like to focus my remarks specifically on the questions that were proposed in the Federal Register in August and try to provide the views of my law firm, and I'll emphasize that these are not necessarily the views of our clients on the problem facing the patent office right now.

Seems to me that it would be unfair to make all biotechnology applicants pay for a problem that's the result of a handful of patent applications. And so at the risk of offending those companies that are filing the patent applications, I think that the cost should be allocated based on the patent office resources that are consumed by these patent applications with large numbers of sequence.

I have some personal experience in looking at this issue. I was involved in filing some patent applications for the National Institutes of Health and became rather notorious with thousands of EST sequences, and I'm not saying that I was -- I was very surprised that

we did not get a thousand way restriction requirement back from the patent office.

It seems that the patent office when faced with a large number of sequences that are related only in a general way, already has a tool for -- for dealing with collecting fees commensurate with the amount of work required for examination.

Another alternative that I could suggest, aside from restriction requirements, would be a surcharge if considerable resources are going to be consumed by a -- for examination of a particular patent application. A surcharge could be imposed, but I think only after giving the applicant an opportunity to amend the claims.

So what I would envision is that the examiner takes a first cut at looking at the claims, looking at that amount of searching that will be required, coming up with a total amount and an explanation of why considerable searching would be required and then giving the applicant an opportunity either to amend the claims or to pay the additional search fee.

I have been surprised, again, with some patent applications I filed where a claim that we really don't expect to get is thrown in for sake of completeness that says, for example, an 8 MER or an 18 MER or 15 MER of the full length sequence that we really want to patent. My understanding is that considerable amount of search resources can be used looking for those short sequences when ultimately those have only an ancillary importance to

the applicant and if given the choice between paying a fee to have a massive search conducted or forgoing the claims for the short sequences, I think a lot of applicants would forgo that broader search. So the search could be attenuated even before it started by giving the applicant the opportunity to amend.

Echoing another approach the patent office could take -- and it's already been suggested -- is to keep the database size down by eliminating the need to submit known sequences and unclaimed sequences, in other words, to allow reference, for example, to a Genbank sequence instead of submitting that sequencing electronically once again and

providing multiple copies in the patent office database.

And again the requirement that sequences of 10 bases or four amino acids be submitted is a burden to applicants and I think, again, clutters up the database.

As to a long-term solution, I've given a lot of thought to patenting of naturally occurring sequences, and I concluded that the current patent system is not well suited for handling naturally occurring sequences that are cranked out in the thousands or tens of thousands at a time. I know there has been concern expressed by scientists working in a field that the rewards of a patent for -- are not necessarily commensurate with the value provided when a genomics company is allowed to get a patent on a sequence based on a very minimal utility.

So one suggestion that I have espoused in recent years is a registration system similar to copyright

for providing some sort of protection for sequences that come out of massive sequencing efforts, naturally occurring sequences. Perhaps that registration system could provide only a defensive right, again, subsequent patents and could also provide perhaps a transferable exclusion of that sequence from the prior art. So that when a utility is determined -- that fits into the more traditional categories that one would anticipate for a patent such as the use of a particular molecule for treating a particular disease or a particular diagnostic application, the prior -- the exclusion from prior art would provide the ability to subsequently patent that sequence for its biological use or for a use related to the protein encoded thereby and the prior publication, the publication of that sequence subsequent to its discovery and registration but prior to the patent focusing on the biological utility would not provide a barrier to getting more traditional patent protection. So those are my thoughts. If there any questions, I'd be happy to entertain them.

MR. LEHMAN: Thank you. You heard me talking about the European database with the previous witness, and that sounds like that's sort of the direction of what you were talking about there.

MR. ISRAELSEN: It's a similar idea, yes.

MR. LEHMAN: Of course, one of the -- that is more of a copyright-based approach, and that doesn't give you any exclusivity vis-a-vis independent creation of the same database.

69

MR. ISRAELSEN: That's correct.

MR. LEHMAN: Should it?

MR. ISRAELSEN: That's a -- that's a loaded question, and I'm probably not the best person to answer that. With my proposal for a registration system for DNA sequences that are naturally occurring, it would not give exclusive rights. It would not give the right to exclude others but would help preserve the option of obtaining a patent at a subsequent time.

MR. LEHMAN: Okay. Are there any other questions?

MR. GOFFNEY: Yeah, I wanted to ask him whether he had any criteria when a search -- a surcharge might be required.

MR. ISRAELSEN: I don't know the economics, but I would expect that it would be imposed only on a very small percentage of applications so that an application containing -- just to pick a number -- 50 sequences would not incur the surcharge.

MR. LEHMAN: Are there any other questions?

Thank you very much.

MR. ISRAELSEN: Thank you.

MR. LEHMAN: We appreciate it. I want to thank the -- particularly the practitioners who came and -- oh, we have one more.

Amy Hamilton.

MS. HAMILTON: Good morning. My name is Amy Hamilton. I'm from Eli Lilly & Company. My views do

70

represent those of the company. Our address is Lilly Corporate Center, Indianapolis, Indiana 46285.

I hadn't planned on speaking today until I sat there and heard nothing but restriction requirements, and then I was thankful to hear Mr. Israelsen mention that because I really think that's incumbent upon the PTO to take applications where there's no overriding, nontrivial

utility to unify the sequences that clearly the law requires that they be treated as separate inventions. And then when that is done, GACT, and its ramifications will, I think, cause these companies to properly file these as separate applications.

I would also recommend some kind of a surcharge for a huge number of sequences. We have thought about, again, some charge appropriately determined by the PTO, something similar to what is done with claims above some reasonable number of sequences. I think 10 is a reasonable number because for a single invention, it's hard to imagine why you would have more than 10 sequences in that application. I'd also like to challenge what was stated in the OT about requiring completeness of examination. 37 CFR Section 105 allows for incomplete examination when there is misjoinder invention or fundamental defects in the application. And I would suggest that cases could be reviewed for a threshold utility before ever making a search, which I think of all things should drastically decrease the amount of computer searching. Then this could be challenged in the courts,

and we could get the utility question resolved in the courts, as we are going to eventually need anyway.

One last thing, 1800 examiners continue to make rejections based on potential methods of making genes and proteins. This form of rejection has been explicitly stated as improper by the CAFC in a series of cases ending in in re: Bell and in re: Dual. This is prolonging examination and taking up valuable examiner time when there is no prior art chemical structure to be used in the rejection, and I would suggest that valuable examiner time could be saved by eliminating this type of rejection.

That's all I have. Thank you.

MR. LEHMAN: Thank you very much. Did you give us your address where you were located?

MS. HAMILTON: I did. I will again. Eli Lilly & Company, Lilly Corporate Center, Indianapolis, Indiana.

MR. LEHMAN: You came all the way out.

MS. HAMILTON: I wanted to come to the first one.

MR. LEHMAN: Thank you very much. We really appreciate it.

I believe that completes -- unless someone else signs up, that completes the list of persons who've asked to testify. Does anybody else like to --

MR. O'HARA: I have maybe one comment. I didn't sign up. My name is Patrick O'Hara. I'm from a company called Zymo Genetics, and that's Z-y-m-o Genetics

in Seattle, Washington, and we're a customer of Incyte's.

And I just wanted to make a reality check on the amount of time that you're thinking of that it's going to take to do the searches because we do quite a of searches ourselves, and I think that you're talking about having 200,000 sequences that you have to search and it's going to cost about $24 million and take two years and nine staff years to deal with it, and I think that Randy said they can do 6,000 a day using blast as an algorithm, and the question, I guess, is blast sufficient for the kind of things that you have to do?

The person from Peterson & Associates said that very sensitive kinds of searches may not really be necessary for the kinds of obviousness criteria that you have to look at. So it seems like what might be possible to do would be that the PTO simply develop a very specific criteria of what it is that is invent- -- invent- -- sorry, invented with regard to the ESTs, and we at Zymo Genetics do about -- we don't -- we don't do 6,000 a day but we do about 1500 a day. And so if you are able to eliminate 90 percent of your problem, instead of being 24 million would be 2.4 million, instead of two years, would be 4.8 months and at about 1500 a day, that would kind of do it, and I think that with blast and related algorithms, plus some post processing of the results, you'd be able to come up with a very nice output that would pretty much automatically determine what it is that you're looking at in most of these EST applications. And I think it would

decrease the number of staff years that you're talking about as well. So while with regard to the kinds of

algorithms that you would have to use, ESTs are parts of genes, and there are specific issues with ESTs that mean that certain kind of algorithms have certain kinds of problems. So, for instance, there's partial overlaps that will give you very different scores than complete overlaps. There are certain issues with regard to accuracy in the database and with regard to the ESTs, and those can cause drastic differences in scores, but, however, those things can all be handled pretty much automatically at a rate of at least 1500 a day.

So it seems to me those problems should be manageable by developing criteria very specifically, and then perhaps contract out the searching so that the PTO with the many kinds of searches or your type with 10 sequences or so, but when you're talking about those mass ESTs, contract it out, develop the criteria and contract it out.

MR. LEHMAN: Thank you very much. I appreciate that.

Are there any other people who wish to be heard? If not, again, I want to thank the people who took the time to come here, particularly individual practitioners who obviously could be doing other things, and note that we're having one further hearing on this. We came here to San Diego because this is a center of the biotech industry. We wanted to make it easier for people

to come here as opposed to come all the way to Washington, but we will be having a hearing on April 23rd in Arlington for anybody who has some comments between now and then wants to get on the plane and come to Washington, but of course we're also happy to receive any further written comments, and we would appreciate it if those would be submitted by April 23rd.

A transcript of this hearing will be made available as soon as it's ready as well as all of the comments received. And we think that's going to be on or about May 13th, 1996. And it will be available for inspection in Room 520 of Crystal Park One, 2011 Crystal Drive in Arlington, Virginia. And it will also be

available on the Internet through a nonfile transfer protocol. The address is ftp.uspto.gov.

All written comments and oral comments made here today will be taken into consideration before implementing policies relating to this issue. And any written comments that are received after the 23rd can't be assured of consideration. I have to say that we really try hard to give people a chance to have input, and then after we're six months into a new policy and somebody says, "Well, you didn't consider this or that," we have not yet had a patent application that gets us access to impressions, so we really need to hear from people if we're going to take their views into consideration.

So I'd like to remind everybody that another hearing will be heard on April 23rd starting at 9:00 a.m., and that will be in the commissioners conference room, Suite 520, Crystal Park One, 2011 Crystal Drive, Arlington, Virginia. And that concludes the hearing. And I thank everybody that participated and shared their views with us.

UNITED STATES DEPARTMENT OF COMMERCE

PATENT AND TRADEMARK OFFICE

- x
:
PUBLIC HEARING ON PATENTING OF :
NUCLEIC ACID SEQUENCES :
:
- x

Crystal Park Two
2121 Crystal Drive
Commissioner's Conference Room
Suite 912
Arlington, Virginia

Tuesday, April 23, 1996

The hearing in the above-entitled matter, commenced, pursuant to notice, at 9:00 a.m.

B E F O R E:

BRUCE A. LEHMAN
Assistant Secretary of Commerce and
Commissioner of Patents and Trademarks

LAWRENCE J. GOFFNEY
Acting Deputy Assistant Secretary of
Commerce and Acting Deputy Commissioner
of Patents and Trademarks

EDWARD R. KAZENSKE
Deputy Assistant Commissioner for Patents

STEPHEN G. KUNIN
Deputy Assistant Commissioner for Patent
Policy and Projects

NANCY J. LINCK
Solicitor

C O N T E N T S

ORAL TESTIMONY BY: PAGE

P R O C E E D I N G S

MR. LEHMAN: Good morning.

My name is Bruce A. Lehman. I am the Assistant Secretary of Commerce and the Commissioner of Patents and Trademarks.

Joining me at this Hearing today are Lawrence J. Goffney, to my immediate left, Acting Deputy Assistant Secretary of Commerce and Acting Deputy Commissioner of Patents and Trademarks; Edward R. Kazenske, Deputy Assistant Commissioner for Patents, on my immediate right; Stephen Kunin, Deputy Assistant Commissioner for Patent Policy and Projects, on my far left; and Nancy Linck, the Solicitor of the Patent and Trademark Office.

This is a hearing to receive public comment on a serious problem that is currently facing the Patent and Trademark Office related to patent protection for nucleic acid sequences. The public was invited to comment on this issue in a notice that we published in the Congressional Record on March 12, 1996.

For over a decade, the PTO has been examining and granting patents to claims reciting nucleic acid sequences. Scientific and technological advances have permitted rapid identification of large numbers of genes or gene fragments. The ease of utilizing automated techniques for sequencing nucleic acid fragments has resulted in the filing of a growing, although still relatively small number, of patent applications, each of which claim thousands of nucleic acid sequences. Statistics reveal that the number of these applications is growing and based on the number of organisms and genes still to be discovered, such growth will continue for the near

future.

In Fiscal Year 1991, the Scientific and Technical Information Center of the PTO searched about 4,000 sequences. In Fiscal Year 95, they searched about 22,000 sequences. Currently, we have over 200,000 sequences claimed in at least 70 patent applications awaiting search and examination.

Our estimates show that the search of 100 sequences requires about 15 hours of computing time but the evaluation of the search results for those 100 sequences requires about 65 hours of examiner time. The PTO currently has two massively parallel processor computers and could run the searches in about two years, with the computers running twenty-four hours a day, seven days a week. To examine this relatively small number of patent applications only with respect to the prior art, however, would require over 90 senior-level staff years. Thus, in order to process these applications, the entire staff of the Biotechnology Patent Examining Group 1800 would have to work for more than nine months exclusively on these applications.

These applications present a challenge to the PTO and we need help and suggestions on how we can address this problem. The United States is a leader in the rapidly growing field of biotechnology, which is a growth industry important to the economic health of this country.

The PTO has taken a very active role in working with its customers to simplify policies and procedures in ways that encourage and promote the growth of this industry. We are committed to improving the responsiveness of the PTO to its customers and to more effectively address the needs of the industry.

We must find ways to search and examine the pending applications and provide these applicants with the appropriate patent protection

for their inventions without creating an imbalance in the appropriation of the resources within and among the technologies and Patent Examining Groups. The policies established must permit the timely and thorough examination of all applications which require the same resources for completion.

We are currently working in partnership with the applicants of these applications in order to explore innovative mechanisms and to accomplish the required work in processing the applications. We appreciate the time each person who is here today has taken to attend the hearing and to provide us with your input into the solutions to these problems.

A transcript of the hearing will be prepared and will be made available for purchase by the public approximately 10 days after this hearing. Copies will also be available directly for purchase from the stenographer.

The name of the stenographer service today is Miller Reporting Company. Their telephone number is (202) 546-6666. That is Miller Reporting, (202) 546-6666. I assume you can probably talk to the stenographer here too about that.

We have received eleven written comments and seven requests to appear orally this morning. However, any persons who wish to speak and who have not previously informed us of their desire to testify are encouraged to add their names to the list located at the credenza at the rear of the room.

In order to permit all persons requesting to appear orally, including those people that may be signing up today to present testimony, we would request that each speaker limit their comments to 15 minutes. You don't have to take your full 15 minutes keep in mind.

Those persons who wish to provide

additional comments must submit their comments to us in writing no later than April-- well, that is today, April 23rd-- no later than today. The speakers have been listed in the order in which the requests were received by us.

You may also pick up at the table at the rear of the room copies of the "Official Gazette", publication of the Notice of the Hearings, and the Request for Comments on the issues relating to patent protection for nucleic acid sequences.

When you present your comments, we would request that you please give your name and address and tell us whether these comments are you own, or whether they are those of your law firm or company, or whether you represent an organization and are presenting comments on their behalf.

We would appreciate it also if the comments could be limited to the questions that were presented in the Federal Register publication of March 12, 1996.

The first speaker that we have listed is Michael Langan, but I was informed that he doesn't appear to be here yet.

Is Michael Langan here?

(No response.)

MR. LEHMAN: If not, is Margaret Smith here?

MS. SMITH: Yes.

MR. LEHMAN: Why don't you come forward, Ms. Smith?

MS. SMITH: My name is Margaret Smith and I represent Genetics Commuter Group, GCG. I am the chief operating officer and one of the owners and founders of GCG.

The mission of GCG is to serve biologists by discovering, implementing, publishing and supporting algorithms in the area of sequence analysis. This is the only thing that we do. This is our main focus.

We also think that standards are very

important, and that is one of the arguments that I would like to use as a basis for my comments. And that is that what standards, such as GCG, which is widely used for analysis and searching, also as a central management tool for public databases, that are installed locally, built on standards using Fortran, C motif, runs on standard platform, such as Sun, SGI and Digital, and also incorporates standards such as Fass Day, Smith Waterman, Blast, SRS, the public databases like GenBank, PIR, Swiss Proit, EMBL, other important files like Rebase and Procite, as well as standard comparison tables.

The software from our company is also a standard on which other people build tools. Within institutions there are modifications so that longer sequences can be used or that searches can be run in a different manner or a set of searches can be run in a certain way.

The GCG software has also been a platform upon which other tools have been built.

For example, in Europe there is something called extended GCG, which is another package built on top where there are modifications and extra programs.

Compugen has built a strategy. Compugen, which is another company, has taken the basics and looked at the problems of high throughput and how to speed up searches and therefore taken the problems that are probably very similar to the Patent Office that are faced by the pharmaceutical companies and other high throughput laboratories where many sequences must be analyzed.

Also, based upon our standards, our standard package, people are building interfaces so that users, scientists, can access just the programs of interest.

The European patent office, for instance, uses our software and has a set menu that their examiners use and therefore they are limiting and

ast using the part of the package that is most appropriate for them.

So, the basis of the argument is that there is a set of standards and we have incorporated many of the standards and become a standard.

I would like to specifically address the issues of the search strategy. As I mentioned before, the pharmaceutical companies face this issue, and there is a proposal of a filtering system. If you have a lot of sequences to go through, that you can first, I will give the analogy, it's like a course grain filter that is very quick. Things move through it very quickly, but you will catch the most obvious.

An example might be blast, the blast algorithm.

Then moving on to things that are more particular, the next part of the filter, would be something along the lines of Smith Waterman or Fass Day perhaps. And the next level of filter could be something like frame search, which is a program that allows you to take a protein sequence and search against a nucleic acid database or vice versa.

I think an important part of this filter system is to have machine readable output. There are standards. The tools to make up the filter are there, but I think the part where it's new research is needed -- not necessarily research, but new work is needed, is to look at it in terms of machine readable output and also normalization, so that there can be an acceptable level of identity before moving on to the next step in the filter. An acceptable level of identity is recognized.

I know that some of the issues seem to deal with ambiguity within sequences and there are also some other programs that can deal with ambiguity within sequences, such as Profile Search, which allows you to indicate regions of high

ambiguity or regions of high stringency before the search is done. There are also pattern matching searches that can be done.

There is really probably a circumscribed set of tools that can be placed in different slots of the filter, depending upon what the problem is. But I think the common theme is machine readable output, as well as normalization of scores.

There is work being done on this already, but I think that this high volume of data that is coming, even though we projected that it would come, everybody, not just GCG, projected that it would come, we are still caught off-base by how quickly it has come.

So, work is being done by us and by others on machine readable output, as well as normalization scores, and I think a lot of that work has been done already.

I would also like to suggest that based on the last ten years of experience that the Patent Office has had with sequence patents, perhaps guidelines can be provided for customers to provide a better set of prior art searches results that the applicants have done and can be submitted with the application. This would move some of the work, and perhaps more of it, off to the applicant to be brought in as prior art information.

An example might be maybe the type of search and parameters that were used, the date of the search, the date of set and the results in a machine readable form and then the examiner can further refine this.

I would also like to point out that the PTO and GCG actually share a large set of customers. And these customers, for example, if you look at the people that have many of the companies that have a high number of patents, such as Tequita, Immunex, Hoffman-LaRoche, Ciba-Giby, Santori, Setas, Eli Lilly, those are shared

customers, as well as NIH, University of Washington, University of Texas, New York State University. These are also public institutions that have a high number of patents in this area.

Therefore those people already have many of these tools, and I think that if they don't, there can also be public access to these tools. But there is an incentive to develop new tools that can be supported within this community and these new tools would be in the area of automatic annotation and normalization.

I would also like to state that the strength of the standard tools, which are well supported through GCG software, as well as the strength of Compugen, which supports high throughput solutions, they are working with pharmaceutical companies and addressing many of these issues already with the pharmaceutical companies on how to work with many sequences.

I would also like to add that GCG has a long-term working relationship with public database providers. This is also a key point in that keeping the data up to date and searching against that data so that searches against that data are at the appropriate time, have the appropriate data for when that application is appropriate.

Also GCG has relationships with other software developers and vendors if new algorithms, perhaps in an adjacent area, are also required.

I think that is it.

MR. LEHMAN: Is it your view that apparently we are not using these tools right now that you are describing in our searches and that if we were to do so, we would very substantially reduce these processing times that we are talking about.

MS. SMITH: I think that some of the tools are being used. Like, for instance, the Smith Waterman. It is my understanding that that is being used. It's hard to say what exactly the

tools. The tools I think are interchangeable depending upon the problem that you are searching.

I am making the argument that it's a standard set of tools, that if it can be provided where you can pick and choose the appropriate filter to slip in to the grid that you are creating of filters and not have to go out and look for the new one outside and try to figure how to work that into the solution.

So, I think you have many of the tools already. It's working them into a filtering system and also the machine readable output and normalization.

MR. LEHMAN: So, in theory, if we are using the most efficient filter, the quickest filter, that we would be likely to catch in enough cases a prior art that would invalidate the patent application or the claims that we could just throw that one out and not have to go further.

Is that the point, that it would safe us a lot of time?

MS. SMITH: Right. Right.

MR. LEHMAN: The difficulty is though if one assumes that these patent applications will be more and more and more eloquent and won't be able to be thrown off that easily, we will still be stuck with the problem of having to do a more comprehensive search and have finer filters.

MS. SMITH: Right. And it's the ability to slip in those finer filters at an appropriate spot that I think is good about working with standards.

The advantage of the standards is that a large community has tested them and they have kind of been through the mill, so to speak, in that errors have been detected and been fixed and people respect the results.

MR. LEHMAN: Now your company makes the software --

MS. SMITH: Correct.

MR. LEHMAN: -- basically that is used.

I assume the primary purpose of that software isn't really to find the prior art in a patent search, it's to aid in the scientific research itself.

MS. SMITH: Our software is used for many different things, but it's really just basic sequence analysis. So, if you can represent sequences as text, and in the case of DNA you have four nucleotides and proteins, 20 amino acids, that can each be represented by a single letter, our software is basically a large text processor.

So, we are continually adding tools that have been recognized as standards or have come up through much testing, and many of these tools have come from public institutions.

So, it's a solution which you can build around. So, it's a core that will allow you to slip in new tools. In fact, that is what some people do at our customer institutions. They will build new tools on top. Because we don't know what everybody needs. So we try and build a common set based on standards that have been tested and in a sense approved by public opinion and then people build things on top of it because we provide the source code and allow that.

MR. LEHMAN: Any other questions?

MR. KUNIN: I would like to ask -- is this mike on?

MR. LEHMAN: Yes.

MR. KUNIN: The presentation that you made, I didn't quite understand the full impact of what you were saying, seemed to emphasize a substantial amount of front end processing with a discussion of machine readable output.

One of the components that we have indicated here as a significant problem of ours is the post processing. And that is to say that in many instances our problem is not in identifying

hits, but our problem is having very large numbers of hits and then having a very difficult time in using Examiners of having to sort through the outcomes and make the actual patentability determinations.

Do you have any recommendations in terms of taking the machine readable output and doing some post processing that would help take those results and put them in a more orderly way to help the examiner on the back end.

MS. SMITH: I think it would depend upon what level of identity you are willing to accept or not accept.

I am making an assumption that the large number of sequences are due to much ambiguity in an original sequence being allowed, but maybe that is a wrong assumption.

But if there is a lot, if a lot of ambiguity is allowed, then I think it's going to be very hard to attack this.

But if the level of identity that you want can be addressed and put at a certain level, then I think the machine readable output is very good at going through and finding if you attain that level of identity, and then only bringing to the examiner the pertinent information that is at the appropriate level of identity.

MR. LEHMAN: Anybody else?

(No response.)

MR. LEHMAN: If not, thank you very much.

MS. SMITH: Thank you.

MR. LEHMAN: Next I would like to ask Eli Mintz, please, if he is here, to come forward.

MR. MINTZ: Good morning.

My name is Eli Mintz. I am the CO of Compugen and speaking on Compugen's behalf. Compugen manufactures hardware for accelerating sequence searches, homology and similarity

searches. I am also speaking here today on behalf of Silicon Graphics International, SGI, and the points I am going to talk about have received their approval.

I would like to raise four points that I think should help the Patent Office to solve some of its problems. Some of these points are the same as Maggie Smith has raised.

The four points are:

(1) Cost effective scalable hardware.

(2) Filtering strategy for searches.

(3) Automating work flow based on rule based systems.

(4) Distributed prefiltering at U.S. PTO customer sites.

I would like to speak briefly about each point.

We think that cost effective scalable hardware solution can be built for doing the searching that U.S. PTO requires. A basic building block would cost $10 per hour and would enable exhaustive searching of 100 sequences each 500 base periods long against a database with sizes one billion base periods in less than 15 hours.

So, relative to the numbers published in the announcement, it's about an improvement of a factor of ten in cost effectiveness.

The chronology that Silicon Graphics has, for example, on how hardware is based on and the GCG software that is used to integrate all this, the technology is all the time evolving and we believe that in the years to come there will be still substantial advancement and the cost of searching will go down.

So, I really don't think that searching will be that much of a problem in the future. I am talking here about the rigorous searches, such as Smith Waterman.

Going back to the filtering strategy, I think it is an important strategy and will bear

fruit say into the future because eventually everything will be prior art. Then the less rigorous searches will quickly fish out results. So it wouldn't make sense to start out with the most rigorous or the more sensitive algorithm right away, but it would be a good idea to prescreen the results, to prescreen the applications, with an algorithm such as Blast that is very quick, not rigorous, but will pick up a lot of homologies.

Let's say in the next few years this may not be the case, but four or five years down the road I am quite confident that most of the filtering will be done at that stage.

The third point, which I think is the one where the U.S. PTO can save a lot is automating the work flow.

Recently there has been some work done by Chris Sanders' group at EBI, the European Bioinformatics Institute, and what they have done, I believe, is a proof of concept of what can be done. Let me just read this out.

The system that they built is called Ginquiz. It is an integrated system for large scale biological sequence analysis that goes from putting sequence to biochemical function using a variety of search and analysis methods and up to date protein and DNA databases.

Applying an expert system module to the results of the different methods, Ginquiz creates a compact summary of findings. It focuses on deriving a predicted put in function based on the available evidence, including the evaluation of a similarity to the closest homologue in the data base, identical, clear, tentative, homoginal.

The analysis used everything that can possibly be extracted from the databases, including three dimensional models by homology when the structure can be reliably calculated.

Ginquiz consists of four modules; the

database update, the search system, the interpretation module and a visualization and browsing system. The models are driven by programs and a front end program for visualization based on a WWW browser is available.

The principal design requirement is the complete automation of all repetitive actions, repetitive subset dates, efficient segment similarity searches, the automated evaluation and interpretation of the results using expert knowledge quoted in rules.

This system has actually been run on a 64 processor Silicon Graphics power challenger ray and it has been used to analyze 6,000 protein sequences from the genome yeast and they plan to analyze the whole yeast genome when it will be available probably later this month.

I think this is a proof of concept. The problem they are trying to solve is more difficult than the problem that the U.S. PTO faces because they want to predict function of a sequence, not just to determine if it is prior art or not. And therefore I think that the technology is available today to allow the U.S. PTO to substantially decrease the Examiner time that it requires and use such automated tools.

The fourth point is, as I said, distributed pre-filtering at U.S. PTO customer sites.

I think the right way to go is to move most of the work to the sites of the people that send in the applications. Because if the U.S. PTO has a defined set of actions that it takes on each application, it would make sense to have the people sending in the applications do this before having sent in the application. And, in my opinion, this can substantially lower the amount of sequences that reach the U.S. PTO, and when they do reach the U.S. PTO, there will already be behind them substantial work done and maybe some of the work

will not have to be repeated.

Especially if, let's say that third parties take the definitions supplied by the U.S. PTO and industry group, it doesn't matter, and build on top of them some software that can be sold or provided to the community, and the U.S. PTO will know that if you use this software, the results are in a format that is acceptable to it, then a lot of the work can be moved from a centralized location towards the people sending in the applications.

That is basically it.

MR. LEHMAN: I take it your assumption is that if we moved this prefiltering out to the applicant site, that they would be able to do it sufficiently rapidly that they would be able to get it into us quickly.

Because, of course, one of the difficulties is if the invention should end up being disclosed in some way, and keep in mind a lot of this research is being done in governmental institutions and some that aren't filing patent applications and maybe publishing the work, that if they are spending a lot of time doing this presearch themselves before they get to us, they would take a risk of having the invention ultimately go into the public domain, certainly in other countries, if not here.

So, would this be able to be done quickly enough, do you think, so that people would not lose an advantage of giving everything they have to the PTO right away?

MR. MINTZ: I believe so. Because the problem that each applicant faces is small relative to the problem that the U.S. PTO faces. Because you have to analyze all the sequences and the applicants have only to analyze their specific sequences. I think this can be a pretty quick process.

Most of the sites that I am aware of can

easily handle just things in house without any problems. They have the capabilities, they have the hardware, they have the software. But they have the building blocks of the software. There needs to be a definition of what exactly should be done with the sequence before sending it into the U.S. PTO.

MR. LEHMAN: Does anybody else have any questions?

(No response.)

MR. LEHMAN: If not, thank you very much.

Did you come all the way from Israel for this?

MR. MINTZ: Yes.

MR. LEHMAN: Next on our list is Hollie Baker.

MS. BAKER: My name is Hollie Baker. I am a senior partner at the law firm of Hale and Dorr in Boston, Massachusetts. I am here on behalf of the American Bar Association, the Intellectual Property Section, where I am Chair of the section's biotechnology committee.

On behalf of the section, I want to thank the Patent Office for having these hearings. We understand how much hard work and time has gone into the hearings, particularly on the reverse side in working with the committee members who have devoted their time and effort in reading the notice and providing some comments to us and preparing our written report which we have submitted to the Patent Office.

We also understand the enormous amount of time that is going to be involved in finding solutions to these issues on nucleotide sequencing.

It wasn't that long ago that I sat in the Patent Office Conference Room with other patent practitioners and with patent examiners where we were first exploring the issue of submitting nucleotide sequences at the first instance. And at

my former law firm, we were one of the test beds for evaluating the patent program.

Technology has moved a long way since that time. Because at the time we were filing patent applications primarily devoted to sequencing of proteins, or perhaps particular sequences with some modifications to make some second generation drugs, vaccines, particular probes for diagnostic uses, and now we are sequencing genomes.

So, just as the technology has moved forward in biotechnology, I think the computer industry has moved forward significantly.

That goes into the first comment that we would like to make.

The Patent Office listed three issues that they wanted to have addressed, and I would like to address the last issue first, which does move into the first issue.

Our section recommends that the Patent Office establish a panel of experts on computer searching to work with the Patent Office to evaluate the Patent Office computer searching and analysis of computer results and to make recommendations for improving the quality and lowering the cost of the searches. With the establishment of this panel, the section believes that it can make the appropriate recommendations for improving the Patent Office search capabilities.

Now, since our committee is not an expert on computers, most of us are patent practitioners and are primarily involved with the patent application evaluation and examination process, we consulted after the April 3rd hearing with a computer company, actually it's a consulting company, who had worked with the SEC, for example, in developing their Edgar program, which you may be familiar with.

This was only to make sure that we were

making appropriate recommendations to the Patent Office in our recommendation that they establish a panel of experts to put together something new for the Patent Office searching capabilities.

Everything that has been mentioned by the previous two speakers was included in a report that they prepared for us, and I have encouraged this company to submit the report as written remarks to the panel. I believe they will be doing that today.

One of the things that was not addressed by the previous two speakers, which was a question by the panel, was: How do you analyze this data once you have prescreened it and once you have prefiltered it? How do you put this together?

This company is not a gene sequencing company. What they really encouraged and were rather appalled by is the amount of data that had to be analyzed in these printouts. They really encourage putting together this information as visual data, as an example, so that perhaps in a graph chart form it can be analyzed in a very particular way.

Apparently this is doable. I don't have the expertise to know how this is done. But this is something that I had not thought of, but certainly is currently capable of being done with the data that is being generated.

In addressing the second issue on underwriting the cost, the section does oppose the imposition of higher fees for patent applications containing a large number of nucleotide and amino acid sequences or long nucleotide or amino acid sequences.

The committee is very aware, the section is very aware as well, that the Patent Office is required under 35 United States Code, Section 41-D, to recover the fees on an estimated average cost of its processing services and materials associated with the patent applications.

While we now know that this hearing is directed to address the problem of a particularly large number of sequences or long sequences, our section based on that makes some recommendations under the nucleotides sequence rules.

Current rules were established that nucleotide submissions had to be made if a biotechnology patent application contained as few as 10 nucleotides or three amino acids, which are 12 nucleotides this requirement applies equally to known prior art sequences, sequences used as probes, sequences used in making constricts, even random sequences that have absolutely no other identifying capability other than their nucleotide sequences.

It also, of course, applies to the claimed invention. And although we are not sure, the Patent Office is obviously more aware of this than we are, we believe it may be contributing to an unwieldy database. We think amendment of the nucleotide sequence rules may eliminate some of the clutter that is in this current database.

The section recommends amending the rules governing the submissions to eliminate the requirement that all nucleotides, amino acid sequences and a fragments disclose be submitted and require only those sequences that are claimed to be submitted.

That obviously doesn't get you around the problem that you are having now because apparently all these sequences are being claimed, but it may help the database.

Alternatively, or perhaps additionally, the section recommends amending these rules so that sequence submission will identify new sequences, whether or not those sequences are being claimed, and identify prior art sequences.

If new sequences are then added to claims during prosecution, the applicant could identify

the sequence identification numbers for those sequences and the Examiner could then search those sequences.

This concludes my remarks.

Do you have questions?

MR. LEHMAN: When you said that the rules should be amended to provide that the applicant provide prior art sequences, are you in a sense suggesting that the applicant provide the prior art that we would use to examine the application, that we wouldn't need to go beyond that?

MS. BAKER: Well, most prior art sequences are read already in the public domain somewhere. What we were recommending is that we identify which ones are prior art and perhaps can give a reference as to where those could be located rather than submitting on a computer sequence submission the prior art sequences. Because that is already somewhere in the public domain.

MR. LEHMAN: Thank you very much.

MS. BAKER: Thank you.

MR. LEHMAN: Is Gary Grace here?

MR. PACE: Pace.

MR. LEHMAN: Pace. Sorry. Gary Pace.

MR. PACE: My name is Gary Pace. I am with the Agricultural Biotechnology Research Unit of Ciba Geigy Corporation, which is located in Research Triangle Park, North Carolina.

I come here today as a scientist with 13 years of corporate research experience and 4 years of experience as a patent liaison, including considerable experience with sequence searching. This combination of experience has given me, I think, a reasonable perspective on the problem facing the PTO regarding issues relating to patent protection for nucleic acid sequences.

The PTO is faced with the problem of how to determine obviousness for applications which disclose sequences. Based on the recent demonstrations to the public, it would appear that

a comprehensive structurally based sequence search with extensive reporting of results is being done to examine these cases.

The PTO has predicted that examination of sequences for obviousness would soon swamp the resources of Group 1800, and I submit would also swamp the applicant's ability to underwrite the examination. It is my opinion that this exhaustive approach is not required to adequately examine applications disclosing sequences.

First, I would like to briefly address the established standards for determining obviousness and then, second, apply them to the specific subject of searching sequences.

The Federal Circuit held in, In re Vaeck, that a prima facie case of obviousness was established by showing that the prior art, first, suggests making the claimed invention and, second, reveals that a skilled artisan would have a reasonable expectation of success in attaining the claimed invention. I suggest that this standard can be applied to the problem faced by the PTO with regard to sequences.

How would the standard manifest itself in a search strategy? Because of the degeneracy of the genetic code, nucleic acid sequences which bear as little as 70 percent identity, can still encode the same protein.

The Federal Circuit held in the Deuel and Bell cases that a partial or complete protein sequence, combined with a general method of cloning, does not make obvious a specific DNA coding sequence. Therefore, by searching protein names or sequences, then identifying any DNA sequences disclosed therein, one will have ascertained those sequences which are obvious under this standard.

If the applicant's disclosure recites sequences which are different, then the applicant

sequences would be nonobvious under this standard. Hence, for genes of known proteins, at most only a protein search should be sufficient to examine the application.

It should be kept in mind that sequence searching algorithms will always find some similarity between the subject sequence and those present in a database. This creates what one of my colleagues calls meaningless homology.

The key question, however, is what constitutes meaningful homology. I suggest that meaningful homology for the purposes of examination would be those values above which there is a reasonable expectation of success in obtaining the claimed invention.

This means that there should be some rationale basis, other than numerical homology, for concluding that one sequence is obvious in view of another.

If one searched the databases for sequences with meaningful homology, it would mean, for example, that a larger word size could be used in the search algorithm. This would significantly speed up the search and reduce the quantity of search results.

On another point, excessively long sequences often arise from genomic cloning methods. Such sequences, when disclosed, frequently identify open reading frames or ORFs. A straightforward way of searching databases during the examination would be to conduct a search on an ORF by ORF basis rather than by breaking this large sequence up into arbitrary pieces.

In addition, multiple strategies creating an exhaustive search are likely not needed.

First, GenBank is available as a combined source with EMBL, which in turn has incorporated the contents of other databases. The only other database which is reported to be significantly different is GENESEO, which specializes in

sequences disclosed in published applications or issued patents. Hence, for purposes of prior art searching, only at most two databases need be searched.

Second, in the case of GenBank/EMBL the database is broken up into smaller libraries by species, order or some other taxonomic unit. Searching within specific libraries is much less time consuming than searching the entire database.

Third, the choice to search program must be an educated one. For example, it has been reported that one commonly used program will find similarities between sequences even where is there is not one single identical residue between them. This would be the stereo typical example of meaningless homology. Without choosing carefully the parameters to use in a search, one obtains reams of output which have limited relevance to the question raised.

Lastly, specialized databases are being developed. For example, a database called DBEST has been created which contains more than 200,000 expressed sequence tags from 26 different organisms. Examination of an application disclosing ESTs might be better accomplished through searching this specialized database.

I would also like to reinforce some of the particular suggestions that have been discussed in the biotechnology industry organization community for the last several weeks regarding sequence listings in general.

First, requiring a sequence to be listed simply because it is disclosed and of a certain length leads to the unnecessary listing of linkers and primers. While these linkers and primers may form part of an enabling disclosure, they are rarely part of the claimed invention. I advocate that only those sequences to be claimed should be required to be part of a formal sequence listing.

Second, it often happens that sequences which are disclosed in a specification are already publicly available. For example, if an applicant has compared their claim sequence to a sequence in the prior art, the sequence rules currently require the prior art sequence also to be listed. This prior art sequence need not be listed since it will not be claimed. Each of these suggestions will reduce the resources needed by the PTO and the applicant to process formal listings.

Lastly, as one who has prepared and repaired many sequence listings, I would like to thank the PTO for distributing their program checker. This program has reduced the frequency of error in our submitted listings, for which we are grateful, since some of our listings have exceeded 50, sometimes reaching 100 pages in length.

To conclude, guidelines and rules are required to define an obviousness standard for sequences and to insure the conduct of informed searches. I believe that the guidance and direction provided by the specification can be turned into an effective search strategy, as is done for key word searches of the prior art. By doing so, the search can be limited to certain criteria based on an understanding of the biology of the claimed invention.

How the claim is drafted should also dictate the search strategy. This approach might require specialized skills, but this is exactly the direction that Group 1800 has taken by hiring examiners with these very skills.

The simplest suggestion that I can make is not to be deceived by an application which discloses sequences. The presence of a sequence, even a claimed sequence, does not necessarily mean that a sequence search is required to adequately determine whether the claimed invention is obvious or not. In many cases the use of gene names, protein names, source organisms, et cetera, can

result in an effective search of the prior art for the purposes of examination. This information can generally be found in the background section of the application.

In addition, most applicants already conduct a search of sequence databases prior to filing. By permitting a submission of sequence searches along the lines of an information disclosure statement, the Examiners may thereby be presented with a useful search strategy and useful search results.

The problems identified by the PTO in regard to these issues require solutions. I submit that some solutions can be achieved quickly and guidelines should be published similar to the approach taken in resolving utility issues.

It should also be realized, however, that biotechnology is one of those areas where technology is outpacing the law. Therefore, it is my belief that these issues will need continual study. Consequently, I suggest that the PTO establish a working group to develop solutions to these problems and to monitor development.

I understand that many offers of assistance have already been made and I wish to add mine to this growing list. This is precisely the type of cooperation needed to address these issues now and in the future.

Thank you.

MR. LEHMAN: Thank you.

Your working group would be presumably outside people, would be industry people, that would be called upon?

MR. PACE: Well, my idea was the working group would be composed of PTO personnel, applicants, scientists, inventors, bioinformatics experts, to work on the problem collaboratively and jointly.

MR. LEHMAN: I am not saying that is not

an excellent suggestion, but there is a potential legal problem. We also heard another recommendation for that. The Federal Advisory Committee Act limits our capacity. There is an overall public policy against not having a million different advisory committees to advise government agencies. We have certain limitations sometimes on our capacity to do that. So we would have to see how that affects these suggestions.

Thank you.

Is there anybody else?

(No response.)

MR. LEHMAN: Thank you very much.

Is Michael Fannon here?

MR. FANNON: Good morning.

My name is Mike Fannon. I am the Director of Bioinformatics at Human Genome Sciences.

As many of you may be aware, HGS is responsible for much of the backlog in terms of these large sequence patents. So I am really pleased to be here to be able to work with you this morning. I have often wondered as we put these things together just how the Patent Office was going to work with these. I am pleased to be able to offer my comments here.

I don't claim to have any expertise in the patent specific issues. However, we have developed substantial expertise in searching and analyzing DNA sequences, and I think much of that can apply in this context.

There are a number of issues associated with searching DNA sequences, and in my work with HGS we, in scaling up our sequencing capacity, really undertook this by a baptism of fire. And there is a number of tradeoffs that you can make in terms of how we actually approach the idea of does the sequence exist in the public domain and how we keep track of use of the sequences.

In particular, I will have some comments

about the sequence search algorithms, the methodology by which we determine whether or not a sequence is similar or related to other sequences, the databases themselves and how we choose what we search against, the organization of the results and the interpretation of the results.

What we find is that you can get reams of output, as we have heard from some of the other speakers, as a result of the searching techniques. We have in our context applied various types of database and user interface techniques to help our scientist wade through those and it may very well be appropriate in this context. And then citation searches, doing some of that in an on line fashion and linking those automatically with the search results.

The issues associated with sequence search algorithms really involve a trade off, and the trade off is pretty simply stated as speed versus sensitivity. That is, the more rigorous the calculation, the longer it is going to take in terms of computational capacity.

When we are looking for searches against prior art, we are really looking for in most cases near identical matches to what is known, in which case many of these techniques can be optimized to search for the near or identical matches, and in this case to reduce the number of these misleading or misinformative matches that you can get by extending these algorithms too far and having them find things that really don't exist or don't have a biological context.

In many cases we found it helpful to combine the algorithms, to use the faster algorithm as a pre-screen to sort through the hundreds of thousands of sequences to create a data set that is much smaller to work with and then to use the more rigorous algorithms to analyze that subset.

So, really in many cases these types of

algorithms tend to be a religious issue in the bioinformatics community. But what we found is the practical solution is to use them, as appropriate, to use the faster algorithms with less sensitivity to do the high speed screening and then to analyze the result set that is the output from that.

Now, the composition of the database you search against is also a very significant issue in any problem that involves large quantities of DNA sequences. Certainly we have all seen in this business the rapid growth of the databases, and, again, HGS has been a very great contributor there as well, in that the technology now enables us to discover sequences at a very high rate.

There is also a well-known redundancy inherent in the public databases and any ability that we can put together to reduce the redundancy of the set that we use as the search set is certainly going to reduce the amount of computational effort required, as well as again helping the Examiners avoid chasing false positives, things that look like hits but really are inappropriate matches of one sort or another.

Another way we could think of subsetting the database for the searches is to organize by date and by species.

In the case of Human Genome Sciences, the bulk of our work is in the human domain, so we tend to analyze our data relative first, initially, against other human sequences and then secondarily against relationships, evolutionary relationships, we may pick up with other species.

So by organizing, the other example that comes to mind is by date. If the application has a certain submission date, if we run that application against the sequences that were known as of that date, then much of the work the Examiner does to determine which came first is in a sense already done.

The matches against newly submitted

sequences won't show up in the listing because they weren't in the database in begin with, and by putting the extra effort into the thinking that goes into what do I search against, we find that, I think, on the back end in terms of Examiner time required to evaluate these matches, that that will go down proportionately.

How the results are organized is an issue that HGS faces on a regular basis as well. We have the capacity to run thousands of sequences in a week and we run these various analytical techniques. The output itself from running these searches now becomes the inputs to our database. That is, we use database management techniques to analyze the search results.

I think much of the state-of-the-art prior to the large sequencing efforts at the Genome centers and from companies like HGS and Insight has been involved in really working on the sequence searching technology. What we are finding is that many of the big shops, such as us, take now the results of those analyses, put them into relational database form. Then we can create listings and create queries that allow the Examiners to really identify those things that are the potential trouble spots, that is, the ones that really are hitting high quality matches to known sequences and the ones that require further evaluation, and to bring those to the foreground very quickly through database management techniques.

Of course, reviewing the search results itself is an extremely time-consuming task. So we would suggest that the Patent Office undertake an initiative to develop some tools for reviewing the database of search results that I am talking about.

This is pretty much standard practice with companies like Insight and HGS, Millenium, the ones that do the large scale sequencing, that their

scientists really don't on the first level interact directly with the search results, but they interact with automatically summarized and catalogued search results using database management techniques.

A logical consistent responsive user interface to that database then would be required to enable the Examiners to make very quick decisions, and in many cases to figure out which ones obviously don't qualify, if we can very quickly go through the set and classify these into groups that identify those things that really are novel, because they don't show sufficient sequence similarity, and just to really, again, help use the computer technology to narrow down the work that the Examiner is required to perform.

A very useful technique that has been explored very much in the public domain by the National Center for Biotech Information is to link the search results with on line databases for citations.

Evaluating results requires a good deal of judgment on the part of the Examiner. What we would envision here is something similar to the Blast searching methodology that is available in the public domain that is sponsored by the National Center for Biotech Information, whereby the results of the search is automatically linked then to GenBank. GenBank is cross-referenced with the midline references to, again, reduce the time, be able to come up with an evaluation of the match in a highly integrated manner without doing a separate sort of literature search. I do believe that the technology is available to really help streamline the problem and I certainly can empathize with the problem you are facing here since we deal with large quantities of DNA sequences and it's very challenging. There is no question about it. The problem domain is very technically demanding and we really applaud your efforts in trying to get a handle on this.

I think, in summary, a combination of techniques is appropriate to streamline the process and I think at every stage of the game, starting from the method that we use to do the searches to what characterizes the database, to how we interact with the search results and make some conclusions, you know, can be developed into an integrated system that would substantially streamline the process as I understand it now.

That is all I have.

MR. LEHMAN: Thank you.

Are there any other questions?

(No response.)

MR. LEHMAN: Thank you very much.

MR. LEHMAN: Is Herb Jervis here?

MR. JERVIS: Good morning.

My name is Herb Jervis. I am the Associate Patent Counsel at SmithKline Beecham Corporation in Philadelphia.

SmithKline Beecham spends about a billion dollars a year on research and development and has an R&D workforce of over 4000 developing new medicines. DNA research is a major platform upon which our discovery programs are built.

In addition to our important alliances with Human Genome Sciences and The Institute for Genomic Research, SB also collaborates with over 140 research institutions and companies involved wholly or partly in DNA research. Meaningful intellectual property protection is fundamental to such an investment.

While other speakers, here today and in San Diego last week, have focused on some of the technical solutions to DNA sequence searching problems cited in the PTO's March 12th Notice and illustrated in the demonstration held on April 3rd, I will confine my remarks to some prosecution-based strategies for easing the searching burden outlined by the PTO.

While the temptation may be present in new areas of technology to search for "quick fixes" to prosecution problems by implementing special rules, such actions I think generally should be resisted because of the isolating effect special regulations and legislations have on the future development of the law with respect to a particular area of technology.

The value of an exceedingly large body of precedent (even if it sometimes appears to be irreconcilable) is that it has resulted from the application of the law to particular facts and has yielded a rich tapestry of illustrations illuminated by accompanying reasoning. It is against this background that the next case can properly be viewed.

It appears from my reading of cases such as Amgen v. Chuagi and Fiers v. Sugano that the Court of Appeals for the Federal Circuit is trying very hard not to treat biotechnological inventions very differently. I think it is fair to characterize the Federal Circuit's view as that DNA is a chemical polymer, albeit a complex one, the rules of chemical practice should apply.

Having made the case against special treatment, I will qualify my remarks and say when viewed rationally and carefully, there should be provisions to assist both the PTO and the applicant in the prosecution of patent applications in certain areas of technology. Certainly, drawing requirements and the rules regarding their representations in mechanical cases and rules concerning the representation of color when it is a distinguishing feature in plant patents illustrate the effectiveness of such an approach.

I have been practicing biotech patent law since the early 1980s and it is from that, dare I say historical, perspective that I wish to cast some more remarks today.

It seems just like yesterday that a young

patent examiner named Jim Martinell (who by the way used to have his name spelled out in genetic code words on his office door) and I struggled with the appropriate language for a declaration to support a deposited microorganism, which would be consistent with the requirements of U.S.C. 112 as we then understood them.

I thank you for indulging me in that bit of reminiscence, but I do so to make a point, and that is the cooperation in this area is key.

The PTO, and Group 1800 in particular, have often cooperated in an open and effective manner with the patent bar in respect of many of these problems surrounding biotech inventions, including the deposit declaration practice referred to, rules governing the representation of nucleic acid sequences. The parties even worked together fostering a legislative initiative (103(b)) thought to be necessary to overcome perceived judicial impediments to biotech process patent. I am not sure the world will forgive us for that.

Biotechnology even has its own section of the MPEP (Chapter 2400). These hearings represent the most recent continuation of that important interaction.

Before making a comment or two about prosecution based strategies, let me make one comment about the technical solutions we have heard today.

I think historically extremely effective interactions relating to biotech prosecution were fostered by the so-called "Group 1800 roundtables" sponsored by the PTO and organizations such as AIPLA, BIO and IPO. It seems to me that the technical aspects of the searching problem would be an ideal topic to be addressed in such a forum.

The success of specialized approaches, notwithstanding, it would be prudent to examine current prosecution practice to see if some

solutions are evident before additional specialized examination regulations are contemplated.

Having said that, let me start by recommending against an approach that has been proffered to simplify searching, and that is a strong reliance on a restriction requirement as a limiting approach. Such an approach I think is fundamentally flawed.

First of all, the problems associated with searching sequences in the context of, for example, the applicants' reliance on functional language, as pointed out in the PTO Notice, does not change if there is a single sequence or a hundred sequences.

Secondly, the over zealous restriction practice is already a subject of much concern within the patent bar. I think further encouragement to utilize restriction to solve these sequencing searching problems would make the problem worse.

In a post-GATT world US inventors are severely disadvantaged by having to file and prosecute large numbers of divisionals, not only in terms of fees, but in terms of ultimate loss of patent term.

I realize some would argue that we are now on the same footing as the rest of the world, except that the rest of the world operates on a much less restrictive unity of invention basis.

Turning to a more positive suggestion, I would like to have the following proposal considered; a two-part examination where non-art issues and art issues are treated separately. I realize that 37 C.F.R. 1.105 and the MPEP Section 707 suggest that office actions be complete and not piecemeal, but I would submit that an office action based on an inneffective search is hardly complete.

If such an approach were to be adopted, issues of utility, enablement of a particular claim scope, adequacy of written description,

definiteness of claim language and the like could be resolved before undertaking the search, thus providing a more focused search inquiry.

I realize that such a suggestion may have the potential for protracting prosecution, but again in a post-GATT world there would seem to be sufficient incentive to resolve such issues quickly.

An efficient 112 dialog between the Examiner and applicant should result in a well-drafted and supported claim of definite scope which will lend themselves to more manageable prior art searching. At the time of the art-based examination, I would envision even greater cooperation between the applicant and the Examiner. It is hard for me to imagine that an applicant would go to the time and expense of filling on a sequence without having done some searching.

Nominally, the results of that search appear as one or more references in an Information Disclosure Statement. If it were helpful to the Examining Corps, why not provide the search itself? I stress if it was helpful because, from my perspective, the most disappointing fact to come out of these discussions so far is that there is not a uniform searching procedure employed.

It would seem to me critical that the PTO, and the patent bar, if appropriate, establish some uniform searching guidelines. If that were so, then some of the search burden may, in fact, be shifted to the applicant.

For example, in the two-part process I have outlined above, should the applicant wish to speed up the prosecution of the search, maybe the search then could be provided by the applicant employing the PTO established protocol. The rules currently in place for making applications special or advancing prosecution would appear to provide a framework for such a process.

Finally, it hasn't escaped my notice that my two-part examination proposal may eliminate one of the Group 1800 Examiners' favorite indoor sports, known in the profession as the old 112/103 squeeze. This feat of intellectual legerdemain occurs when an Examiner seeking to reduce the claim scope alleges that the art to which the invention relates is highly unpredictable, thus the claims must be limited to working examples.

Then in the next paragraph when the Examiner is suggesting that the invention is obviously in view of two tenuously related pieces of prior art, all of a sudden the art to which the invention relates becomes predictable, giving rise to a reasonable likelihood of success.

I won't say this two-part examination process will remove this habit entirely, but separating the grounds of rejection temporally may reduce the temptation.

Thank you for the opportunity of speaking today. I remain ready to assist in resolution of this problem.

MR. LEHMAN: If I understand you right, you are suggesting that by having a two-part examination that we would, first, prior in some cases to doing the search, we would deal with the Section 112 and maybe the Section 101 issues before getting to the search to determine obviousness.

Is that correct?

MR. JERVIS: Yes.

I understood one of the problems is in some of the claim language where the claim recites that all sequences hybridizable therewith, and there is no definition of what that is. It makes a search almost impossible if there is even a single sequence in the application, let alone 100.

MR. LEHMAN: In the software area we recently issued guidelines or we sort of moved in the opposite direction because there was a concern in the industry that we were in effect trying to

get rid of work sometimes by, first, tormenting some of our applicants on 112 and 101 issues. And so we sort of revised our procedures in the form of guidelines -- Nancy can correct me if I am wrong about this -- to more integrate the process. So this would be sort of moving in the other direction.

Would it?

MS. LINCK: We certainly have taken the position that a search should be done prior to the entry of any Section 101 rejections. However, in working with the examining core, it really turns out that the 112 issues do need to be addressed. Although they are not to be entered as a rejection, they need to be addressed prior to conducting this search.

So perhaps maybe the order leading into the application and examining the application would be a compromise between actually having two discreet steps.

I don't know. We hadn't thought about this with respect to the biotech area.

MR. JERVIS: If the 112 issues were resolved first, it may resolve the problem of doing the adequate search in terms of timing and still maintaining a priority date as well because that could be postponed to sometime later.

MS. LINCK: I think we found with the software guidelines that it will be required that the 112 issues be looked into early on.

So, that would be consistent with what Mr. Jervis is suggesting.

MR. JERVIS: Thank you.

MR. KUNIN: You indicated that you opposed restriction requirements. You also seemed to try to draw a line between U.S. restriction practice and, as I believe you characterized it, a more liberal international standard of unity of invention.

But isn't it the case that in applying the international unity of invention standard in places like Japan and the European patent office and elsewhere, that for these types of cases they are holding that the sequences lack the unity of invention?

MR. JERVIS: Well, my personal experience, I haven't actually prosecuted the mega types of sequence cases in the European patent office yet. I have never received in the European patent office a 218 waive restriction requirement, I must say, or unit invention requirement. I have gotten those from the U.S. Patent Office.

What I am saying is that I think it's the application of those rules that needs to be tightened a bit.

But, in general, we file many less divisional applications in Europe than we do in the U.S. as a general rule in biotechnology, not just sequence cases.

MR. LEHMAN: But in this particular category of cases, I think it's our understanding that you would have to file 218 individual patent applications in the European Patent Office.

MR. KUNIN: I believe in the European Patent Office they are making you take them one sequence at a time and saying that each sequence is an independent compound.

MR. JERVIS: Not in my experience they haven't so far.

MR. LEHMAN: Any other questions?

(No response.)

MR. LEHMAN: Thank you very much.

MR. SHPUNTOFF: Thank you.

My name is Albert Shpuntoff. That is S-H-P-U-N-T-O-F-F, being the only name so far that isn't easily spellable from pronunciation.

I work as a consultant very closely with Mass Par. I was formerly the education manager and leader of the post implementation bioinformatics

professional services for Mass Par.

Having listened to the comments here and in La Jolla, I found a lot of commonality with some of the things we have done with customers and felt a need perhaps to avoid having it seem like we were blind sighting the people we have worked with at the patent office as incumbent in the massively parallel processing area with new applications, as one of my colleagues presented at La Jolla.

So, first I would like to respond to a couple of things that were said earlier today.

The wide use of a program does not immediately make it a standard for the industry. There is no standard organization at this point that is saying what hardware or software is standard for this industry. There are a number of competitors who claim a standard at this point is aggressive marketing, which should be seen as such.

I am an admirer of aggressive marketing, but I think one should avoid claiming standards until they truly are standards by an appropriate accrediting organization or industry panel.

I have been involved in implementing a number of automated sequence processing schemes at a number of Mass Par's customers and have been active in advising others who are at the point of starting up an operation.

We have seen quite a variety of strategies for dealing with the overwhelming number of sequences that are to be processed in the discovery phase, as well as is there any point in pursuing it to the Patent Office. So I have seen much of the issues from the other side.

No matter how much processing gets increased, the amount of proposed discovery increases at least as fast. So while we propose ways of working smarter and working harder to obtain results quicker, I think it's fair to say

that the workload does tend to increase as the computing power increases. And as the availability of attractive computing to do bioinformatics in the industry has increaseed, so certainly has the number of novel results to be presented to the Patent Office.

We have customers who have chosen to do annotation searches on a relational database version of the standard databases to produce a custom version of the database to be searched.

Because it is possible quickly on the Mass Par to produce a searchable database for Smith Waterman searching, we have a number of customers who have chosen to use annotation to restrict the search space and thereby to do more efficient searches, making better use of the parallel time available.

We have a number of sites who have chosen to preprocess databases so as to eliminate potential matches that you have to look at over and over again with different annotations.

There has been a number of mentions of standard software from NCBI that people are recommending be used. One of the products available from them is a product called NRDB, which allows you to take multiple databases and remove identical sequences from additional copies of the database.

There has been a lot of comment at these hearings about the issue of meaningless homology, false hits, large amounts of data coming out of the individual search runs on machines.

We would like to support the idea that many of the false matches are involved with such things as repetitive DNA elements, which are parts of sequences that are to be submitted, such things as aloe repeats.

There are databases of just the aloe repeats which might well be an additional very useful augmentation to the search to help to

identify in perhaps an automated fashion what are the meaningless homology hits during the search process.

In setting up high throughput screening for several pharmaceutical companies this has been sufficiently important to make the operation practical that we have made it the second step in an automated processing. After identifying that the sequence being looked at is of sufficient quality to even proceed with analysis, the next step has been to mask out repetitive DNA and aloe sequences with the idea of providing more high quality hits and reducing the number of false hits that have to be followed through by an Examiner.

So, this has been an important part, I believe, in the La Jolla hearings. The gentleman from Insight spoke about this also being part of their automated sequence processing.

To the degree that the Patent Examiners are spending too much time following hits based on repetitive sequences not key to the art, but which show up as statistical artifacts, it is possible to develop a process which involves masking those prior to running searches.

MR. LEHMAN: Does that have to be done like on an individual basis, each time you have to sort of program?

You call it annotation. Do you have to annotate the program for the particular search that you are doing so that the Examiner would have to become proficient in that search technology or is that something you could standardize?

MR. SHPUNTOFF: Where we have done this and had it contribute, we have used a database which is available at NCBI called the rep base, which is a database of repetitive DNA elements. It is possible to run a relatively simple search. Even Fass Day or Blast are sufficient to identify these repetitive sequences. You don't particularly

care to have the most sensitive search for things that are going to be masked out.

But even if it's just to run an additional search of these to aid the examiner as a general what in this sequence is in the aloe base or repetitive DNA base already to aid in not following with additional searches, those things which are relatively unlikely to be of significance.

In the case that we have automated we actually make this a formal first step and some small percentage of the sequences being searched are almost immediately ruled out as being completely aloes. But for the most part this allows for the output of the first round of searches to be of higher significance by eliminating the most common matches.

Another focus of the work we have been doing has been to post process the outputs of MP search. As we have heard, there is a lot of interest in what can be done in terms of post processing.

Frequently our customers have been taking the outputs and putting them into either relational databases for maintaining sequence databases or feeding them back into lab management systems based on standard database systems.

We are able to parse the reports that are produced using the statistics that are an output of the Smith Waterman technique to provide keys to what further searches might be done, and in a number of the lab settings have been able to choose, based on whether or not we have high matches of a particular sort early during the process, to automatically generate additional searches which are appropriate based on the kind of match that occurred at the beginning.

So, in some settings this has led to the equivalent of a family of paradigm searches which Examiners might use through something like the

existing graphic interface to deal with sort of standard searches based on particular paradigms where requests fall within particular sets of parameters.

I think I will stop with that.

MR. LEHMAN: Thank you.

Steve.

MR. KUNIN: Could you comment on the suggestion that was made in San Diego by the representative from Mass Par with respect to the clustering technique?

There was an indication that perhaps a number of the sequences had a fairly high degree of similarity that they could be clustered together and perhaps treated as a batch.

Do you have any additional comments along those lines.

MR. SHPUNTOFF: I think that what John Burke was talking about is very current research that is being used throughout the community. There will be a number of conferences in which various clustering techniques will be talked about in the very near future with the idea of improving the way with which we do searches.

The databases at present contain much redundancy, many attempts to sequence things, which have resulted in what should be identified as a cluster and for which there could be a consensus gene representation generated through techniques such as clustering and multiple sequence alignment.

The database that was mentioned earlier today, DBEST, is an attempt to try to provide a cleaner database for expressed sequence tags with the attempt being to do the sort of clustering analysis and multiple sequence alignment prior to presenting the database to be searched.

Whether the processing is done within the Patent Office to maintain an indexed and smaller

database, whether the clustering technique is used to help understand the commonality in sequences being provided by an applicant, we feel it's a useful technique which has been parallelized for the machine and one which should be integrated into the processes.

The discussion earlier about word base search and the possibility of using longer word size in doing searches that would be a little bit faster would be relevant to the particular software he was discussing and PT-2 cluster which was developed at the University of Houston and of which he is one of the authors.

MR. LEHMAN: Thank you very much.

Has Michael Langan arrived yet?

(No response.)

MR. LEHMAN: If not, that concludes the witnesses who I have signed up.

Has anyone else signed or is there anyone else who would like to testify at this time?

(No response.)

MR. LEHMAN: If not, that concludes our hearing.

Let me mention that the written comments on the notice of hearings and request for public comment must be submitted by today, I guess, by the end of business today. And a transcript of this hearing will be made available as soon as we can make it available, as well as all the written comments that we have received, and that will be available for inspection and review on or about May 13, 1996 in Room 520 of Crystal Park 1. That is at 2011 Crystal Drive. And it will be available also on the Internet at ftp.uspto.gov. Ftp.uspto.gov.

All written comments and oral comments made here today will be taken into consideration before we implement any policy on this issue. We can't promise though that any comments received after today will be taken into consideration.

I would like to remind those present

today that the next public hearing will be held on May 2nd from 9:00 a.m. to 5:00 p.m. in this room on issues related to patent protection for therapeutic and diagnostic methods. And notice for comments and public hearing on that issue was published in the March 13, 1996 Federal Register, 61 Federal Register 10320 and also in the Official Gazette, 1185 OG 64.

That concludes the hearing today. I would like to thank all the witnesses that came for their help and all the interested parties that came to hear what they had to say.

Thank you.

(Whereupon, at 10:40 a.m. the proceedings adjourned.)